

Aufklärung der Populationsstruktur von  
*Staphylococcus aureus* anhand genomweiter  
Polymorphismen

Von der Fakultät für Lebenswissenschaften  
der Technischen Universität Carolo-Wilhelmina zu Braunschweig  
zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
(Dr. rer. nat.)  
genehmigte  
DISSERTATION

von Kevin Kurt-Spamer  
aus Wernigerode

1. Referent: Privatdozent Dr. Ulrich Nübel

2. Referent: Professor Dr. Michael Steinert

eingereicht am: 21.01.2013

mündliche Prüfung (Disputation) am: 14.05.2013

Druckjahr 2013

## Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

### Tagungsbeiträge

K. Kurt, M. Henkel, JP Rasigade, F. Laurent, S. Ritchie, R. Göring, H. Zemlickova, MJ. Strulens, S. Holtfreter, W. Witte and U. Nübel: Population analysis of a *Staphylococcus aureus* clone ST121 from different lines of infections. - Poster at the 5<sup>th</sup> European Conference on Prokaryotic and Fungal Genomics, September 18-21, 2011 in Göttingen, Germany

K. Kurt, M. Henkel, JP Rasigade, F. Laurent, S. Ritchie, R. Göring, H. Zemlickova, MJ. Strulens, S. Holtfreter, W. Witte and U. Nübel.: Subpopulations of *Staphylococcus aureus* ST121 are associated with distinct clinical pictures. - Presentation at the 63 Jahrestagung der Deutsche Gesellschaft für Hygiene und Mikrobiologie (DGHM) e.V., September 25 to 28, 2011 in Essen, Germany

K. Kurt, M. Henkel, JP Rasigade, F. Laurent, S. Ritchie, R. Göring, H. Zemlickova, MJ. Strulens, S. Holtfreter, W. Witte and U. Nübel.: Subpopulations of *Staphylococcus aureus* ST121 are associated with distinct clinical pictures. - Presentation at the Multidrug Resistant High Risk Clones (MDR-HiRiC): Features, Epidemiology and Detection, 5 to 6 June 2012, Barcelona, Spain

## Summary

*Staphylococcus aureus* is one of the most common human pathogens and causes a wide range of infections, from skin and soft tissue infections (SSTIs) to life threatening diseases. SSTIs predominate in two clinical entities; (i) the superficial infections that affect the upper skin layers (stratum corneum and stratum granulosum), and (ii) deep-seated infections which affect the dermis and subcutaneous tissues. Methicillin-susceptible *S.aureus* (MSSA) complex CC121 is globally distributed, and it is a common cause of both superficial skin infections and deep-seated infections. Commonly, CC121 isolates are susceptible to methicillin. However, recently methicillin resistance had been found among community-associated *S. aureus* CC121 from Cambodia.

In the present study, we investigated the population structure of CC121 by mutation discovery at 115 genetic housekeeping loci from each of 154 isolates, sampled on five continents. In addition, we pyro-sequenced the genomes from ten representative isolates to get deeper insights into the evolution of this clone. The genome-wide SNPs that were ascertained revealed the evolutionary history of CC121, indicating at least six major clades (A to F) within the clonal complex and dating its most recent common ancestor to the pre-antibiotic era. This study was able to show, that distinct phylogenetic lineages within CC121 are associated with specific diseases. The toxin gene complement of CC121 isolates was correlated with their SNP-based phylogeny. Moreover, we found a highly significant association of clinical phenotypes with phylogenetic affiliations, which is unusual for *S. aureus*. Isolates from superficial infections clustered in clade F, whereas isolates from deep-seated infections were disseminated in several clades, but not in clade F.

Our results demonstrate that phylogenetic lineages with distinct clinical properties exist within an *S. aureus* clonal complex, and that genome-wide SNPs serve as powerful discriminatory markers, able to identify these lineages in more details. The comparison of ten CC121 genomes revealed little variation within the core genome but high variation within mobile genetic elements. Nevertheless, all CC121 genomes harboured a 41-kilobase prophage that was dissimilar to *S. aureus* phages sequenced previously. Furthermore, on the basis of genome wide SNPs we could shown that community-associated MRSA and MSSA from Cambodia were extremely closely related, suggesting this MRSA arose in the region.



## Zusammenfassung

*Staphylococcus aureus* gehört zu den bedeutendsten menschlichen Krankheitserregern und kann eine Vielzahl von Infektionen verursachen. Diese können von lokalen Infektionen der Haut und Weichteilgewebe bis zu lebensbedrohlichen systemischen Infektionen reichen. Haut- und Weichteilinfektionen werden vorwiegend durch zwei klinische Bilder mit unterschiedlicher Pathogenese vertreten. Dabei kann zum einen in oberflächliche Infektionen, die ausschließlich die oberen Schichten der Haut innerhalb der Epidermis (*stratum corneum*, *stratum granulosum*) betreffen und zum anderen in tiefgehende Infektionen unterschieden werden. Methicillin-sensibler *Staphylococcus aureus* (MSSA) des klonalen Komplex CC121 ist häufig mit diesen beiden klinischen Bildern assoziiert und weltweit verbreitet. Interessant an CC121 ist, dass dieser erst kürzlich auch erstmals als MRSA in Asien in Erscheinung trat.

In der vorliegenden Arbeit wurde die Populationsstruktur durch Vergleiche von 154 Isolaten des CC121 aus 27 Ländern, verteilt über 5 Kontinente untersucht. Dazu wurden die Isolate nach Mutationen in 115 genetischen Loci unter Verwendung einer spezialisierten, denaturierenden Hochdruckchromatographie-Methode analysiert. Darüber hinaus wurden zehn Isolate, die repräsentativ für die gegebene Populationsstruktur von CC121 sind, mittels 454-Pyrosequenzierung sequenziert. Dadurch war es möglich die Evolution des Klons mit hoher Auflösung zu untersuchen. Ausgehend von der ermittelten Phylogenie des Klons konnten sechs monophyletische Gruppen (A bis F) abgeleitet werden, die im Laufe der Entwicklungsgeschichte von CC121 unabhängig voneinander entstanden sind. Des Weiteren konnte gezeigt werden, dass der Großteil der Populationsstruktur lokaler Natur ist und die Isolate gemäß ihres geographischen Ursprungs clustern.

In dieser Arbeit konnte erstmals gezeigt werden, dass die Anwesenheit der Toxingene (*eta*, *etb*, *lukS/F-PV*) stark mit der phylogenetischen Verwandtschaft der Isolate assoziiert ist, und dass sie bei den Isolaten, für die diese Informationen verfügbar sind, auch weitgehend mit den klinischen Symptomen korreliert. Demnach clustern alle Isolate verantwortlich für oberflächliche Infektionen in Klade F, während alle Isolate verantwortlich für tiefgehende Infektionen sehr divers über die Phylogenie verteilt sind, mit Ausnahme von Klade F. Die Identifikation definierter SNPs ermöglicht es, Linien mit unterschiedlichen klinischen Eigenschaften zu identifizieren und somit das davon ausgehende Gefahrenpotential einzelner Isolate abzuschätzen.

Der Vergleich der vollständigen Genomsequenzen zeigte wenig Variationen im Kerngenom, aber eine hohe Diversität im akzessorischen Genom. Neben einer Vielzahl von identifizierten Prophagen, konnte auch ein bisher unbekannter, 41 kb großer Prophage identifiziert werden:  $\Phi$ SaCC121. Dieser konnte vollständig und hoch konserviert in allen zehn untersuchten Isolate detektiert werden. Weiterhin ermöglichten die populationsgenetischen Analysen basierend auf der Detektion von genomweiten Einzelnukleotidpolymorphismen (SNP) und die genomischen Vergleiche den Nach-

weis, dass eine enge Verwandtschaft zwischen den CC121-MRSA Isolaten und CC121-MSSA Isolaten aus Kambodscha vorherrscht. Die basale Stellung der MSSA-Isolate unterstützte dabei die Vermutung, dass die CC121-MRSA Population in Kambodscha direkt aus der dort vorherrschenden MSSA Population hervorgegangen ist.

# Inhaltsverzeichnis

ABBILDUNGSVERZEICHNIS	viii
TABELLENVERZEICHNIS	ix
<b>I Einleitung</b>	<b>1</b>
<b>1 Einleitung</b>	<b>2</b>
1.1 <i>Staphylococcus aureus</i> . . . . .	2
1.1.1 Die Spezies <i>Staphylococcus aureus</i> . . . . .	2
1.1.2 Verbreitung und Bedeutung von <i>S. aureus</i> als Krankheitserreger	3
1.1.3 Pathogenität und Virulenzeigenschaften . . . . .	4
1.1.4 Antibiotikaresistenzen . . . . .	7
1.1.5 Entwicklung von Methicillin-resistenten <i>Staphylococcus aureus</i>	9
1.1.6 Genomstruktur von <i>S. aureus</i> . . . . .	10
1.1.7 Molekulare Typisierungsmethoden . . . . .	11
1.1.8 Klonale Populationsstruktur . . . . .	12
1.2 Sequenztyp 121 . . . . .	14
1.3 Zielsetzung . . . . .	15
<b>II Material und Methoden</b>	<b>16</b>
<b>2 Material</b>	<b>17</b>
2.1 <i>Staphylococcus aureus</i> Isolate . . . . .	17
2.2 Geräte . . . . .	18
2.3 Chemikalien und Enzyme . . . . .	18
2.4 Nährmedien . . . . .	19
2.5 Kits . . . . .	19
2.6 Primer . . . . .	20
2.7 Software und Internet-Tools . . . . .	20
<b>3 Mikrobiologische Methoden</b>	<b>22</b>
3.1 Bakterienanzucht und Stammhaltung . . . . .	22

3.2	Präparation chromosomaler DNA . . . . .	23
3.2.1	Bestimmung der DNA Konzentration . . . . .	23
3.3	Polymerase-Kettenreaktion . . . . .	24
3.4	Aufreinigung der PCR Produkte . . . . .	25
3.5	Sequenzierungen . . . . .	26
3.6	SNP-Analyse mittels dHPLC . . . . .	29
<b>4</b>	<b>Bioinformatische Auswertung von Sequenzdaten</b>	<b>31</b>
4.1	SNP-Analyse . . . . .	31
4.2	Umgang und Bearbeitung von Genomsequenzen . . . . .	32
4.3	Phylogenetik von CC121 . . . . .	34
4.3.1	Phylogenie mit der <i>Maximum-Likelihood</i> -Methode . . . . .	34
4.3.2	Phylogenie mittels Bayes'sche Analysen . . . . .	35
4.3.3	Datierung und Substitutionsraten mittels BEAST . . . . .	36
4.4	Korrelation zwischen phänotypischen Merkmalen und der Phylogenie	37
4.5	Paup* zur Berechnung des Homoplasie-Indexes (HI) . . . . .	38
4.6	Sequenzevolution $dN/dS$ . . . . .	38
4.7	Genom-Annotation . . . . .	39
<b>III</b>	<b>Ergebnisse</b>	<b>40</b>
<b>5</b>	<b>Ergebnisse</b>	<b>41</b>
5.1	<i>S. aureus</i> -Isolate . . . . .	41
5.2	SNP-Analyse mittels dHPLC . . . . .	41
5.2.1	Sequenzierung der detektierten SNPs . . . . .	42
5.3	Phylogenetik von CC121 . . . . .	44
5.3.1	<i>Minimum-Spanning-Tree</i> . . . . .	44
5.3.2	<i>Maximum-Likelihood</i> -Modell . . . . .	46
5.3.3	Phylogeographie . . . . .	47
5.3.4	Klinischer Phänotyp . . . . .	50
5.3.5	Verteilung der Toxingene . . . . .	52
5.3.6	Assoziation zwischen Phylogenie und Toxingenen sowie klinischem Phänotyp . . . . .	55
5.3.7	Untersuchung zur Sequenzvariation der <i>lukS/F-PV</i> Gene . . . . .	56
5.3.8	Evaluation der <i>spa</i> -Typen . . . . .	59
5.4	Sequenzalignment-basierte Statistiken . . . . .	62
5.4.1	Analysen zur molekularen Uhr mittels BEAST . . . . .	62
5.4.2	Ermittlung von $dN/dS$ . . . . .	62
5.4.3	Analyse zum Homoplasie-Index . . . . .	62
5.4.4	Nukleotid-Diversität ( $\pi$ ) . . . . .	63

5.5	Genomanalysen . . . . .	64
5.5.1	Phylogenie . . . . .	66
5.6	Genomvergleiche . . . . .	68
5.6.1	Pathogenitätsinseln . . . . .	70
5.6.2	Prophagen . . . . .	72
5.6.3	Plasmide . . . . .	76
5.6.4	Andere Virulenzdeterminanten . . . . .	77
5.6.5	SCC <sub>mec</sub> Genkassette . . . . .	78
5.6.6	Analysen zur MRSA-Evolution in CC121 . . . . .	79
<b>IV</b>	<b>Diskussion</b>	<b>81</b>
<b>6</b>	<b>Diskussion</b>	<b>82</b>
6.1	Mikroevolution von CC121 . . . . .	82
6.1.1	Phylogenie von CC121 . . . . .	82
6.1.2	Phylogeographie . . . . .	83
6.1.3	Phylogenetische Relevanz von <i>spa</i> -Typen . . . . .	85
6.1.4	Evolution klinisch unterscheidbarer Subpopulationen . . . . .	86
6.1.5	PVL-Genotypen . . . . .	88
6.1.6	Ursachen für die evolutionäre Divergenz zweier Teilpopulationen . . . . .	89
6.2	Vergleichende Genomanalysen . . . . .	91
6.2.1	<i>S. aureus</i> Pathogenitätsinseln (SaPIs) . . . . .	91
6.2.2	Prophagen . . . . .	92
6.3	Die Bedeutung von CC121-MRSA . . . . .	93
6.3.1	Phylogenetische Beziehung . . . . .	94
6.3.2	Genomvergleiche . . . . .	94
6.3.3	CC121-MRSA . . . . .	96
6.4	Ausblick . . . . .	97
<b>A</b>	<b>Anhang</b>	<b>99</b>
	<b>LITERATURVERZEICHNIS</b>	<b>122</b>
	<b>DANKSAGUNG</b>	<b>139</b>

# Abbildungsverzeichnis

1.1	<i>S. aureus</i> auf Mueller-Hinton-Blutagar . . . . .	3
1.2	<i>S. aureus</i> bei der Besiedlung humaner Epithelzellen der Nase . . . . .	4
1.3	Antibiotika Entwicklung und Ausbildung von Resistenzen . . . . .	8
1.4	MRSA aus Blutkulturen in europäischen Krankenhäusern (2010) . . . .	10
1.5	Populationsstruktur von <i>S. aureus</i> basierend auf MLST-Daten . . . .	12
5.1	Sequenzalignment exemplarisch für Locus <i>au210</i> . . . . .	42
5.2	<i>Minimum-Spanning Tree</i> basierend auf SNP-Daten . . . . .	45
5.3	<i>Maximum-Likelihood</i> -Baum beschreibt die Phylogeographie von CC121	47
5.4	Phylogeographie in einem <i>Minimum-Spanning Tree</i> dargestellt . . . .	49
5.5	Verteilung von klinischen Phänotypen und Toxingenen in <i>Minimum-Spanning Trees</i> . . . . .	51
5.6	Verteilung der Virulenzdeterminanten ETA, ETB und PVL in einem <i>Minimum-Spanning Tree</i> . . . . .	53
5.7	Verteilung der Virulenzdeterminanten <i>eta</i> , <i>etb</i> und <i>lukS/F-PV</i> in einem <i>Maximum-Likelihood</i> -Baum . . . . .	54
5.8	Verteilung von PVL-Genotypen dargestellt im <i>Maximum-Likelihood</i> -Baum . . . . .	58
5.9	<i>Maximum-Likelihood</i> -Baum zeigt die Verteilung der sechs häufigsten <i>spa</i> -Typen . . . . .	60
5.10	<i>Maximum-Likelihood</i> -Baum zeigt die Homoplasien für fünf <i>spa</i> -Typen	61
5.11	<i>Maximum-Likelihood</i> -Baum zeigt phylogenetische Verteilung der Genomsequenzierungen . . . . .	65
5.12	Phylogenie des klonalen Komplexes CC121 basierend auf 1828 SNPs .	67
5.13	Sa1-Prophagen . . . . .	73
5.14	Sa2-Prophagen . . . . .	73
5.15	ΦSaCC121 . . . . .	75
5.16	SCC <i>mec</i> -Kassette . . . . .	79

# Tabellenverzeichnis

1.1	Pathogenitäts- und Virulenzfaktoren . . . . .	5
2.1	Verwendete Geräte . . . . .	18
2.2	Verwendete Chemikalien und Enzyme . . . . .	18
2.3	Verwendete Nährmedien . . . . .	19
2.4	Verwendete Kits . . . . .	19
2.5	Verwendete Software . . . . .	20
2.6	Verwendete Internet-Tools . . . . .	21
5.1	Klade spezifische SNPs . . . . .	44
5.2	BaTS-Analyse zur Phylogenie/Geographie Assoziation . . . . .	48
5.3	BaTS-Werte für die Assoziation zwischen Phylogenie und klinischem Phänotyp . . . . .	55
5.4	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Virulenzgen- Gehalt . . . . .	55
5.5	Sequenzpolymorphismen in der genetischen Determinante für PVL ( <i>lukS/F-PV</i> ) . . . . .	56
5.6	BaTS Analyse zur Phylogenie/ <i>spa</i> -Typ Assoziation . . . . .	59
5.7	Raten und Daten mittels BEAST . . . . .	62
5.8	Genom Statistiken . . . . .	64
5.9	Detektierte Sequenzpolymorphismen im <i>lukS/F-PV</i> -Gen . . . . .	69
5.10	Geprüfte Adhäsions- und Oberflächenprotein codierende Gene . . . . .	77
5.11	Übersicht über die getesteten Virulenzgene . . . . .	77
5.12	Übersicht über getestete Regulons . . . . .	78
A.1	CC121-Isolate . . . . .	100
A.1	CC121-Isolate . . . . .	101
A.1	CC121-Isolate . . . . .	102
A.1	CC121-Isolate . . . . .	103
A.1	CC121-Isolate . . . . .	104
A.2	Identifizierte Polymorphismen . . . . .	105
A.2	Identifizierte Polymorphismen . . . . .	106
A.2	Identifizierte Polymorphismen . . . . .	107
A.2	Identifizierte Polymorphismen . . . . .	108

A.2	Identifizierte Polymorphismen . . . . .	109
A.3	Primer für die Verifizierung von SNPs in den CC121-Isolaten . . . . .	110
A.3	Primer für die Verifizierung von SNPs in den CC121-Isolaten . . . . .	111
A.3	Primer für die Verifizierung von SNPs in den CC121-Isolaten . . . . .	112
A.3	Primer für die Verifizierung von SNPs in den CC121-Isolaten . . . . .	113
A.4	Primer für die Verifizierung von Toxingenen . . . . .	114
A.5	BaTS Analyse zur Phylogenie/ <i>spa</i> -Typ Assoziation . . . . .	115
A.6	BaTS-Analyse zur Phylogenie/Geographie Assoziation . . . . .	115
A.7	BaTS-Werte für die Assoziation zwischen Phylogenie und klinischem Phänotyp . . . . .	115
A.8	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Virulenzgen- Gehalt . . . . .	116
A.9	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und ERY . .	116
A.10	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und PEN . .	116
A.11	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und TET . .	116
A.12	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und SXT . .	116
A.13	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und CLI . .	117
A.14	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und CIP . .	117
A.15	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und OXA . .	117
A.16	BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und FUS . .	117



# Teil I

## Einleitung

# 1. Einleitung

## 1.1 *Staphylococcus aureus*

### 1.1.1 Die Spezies *Staphylococcus aureus*

Als Robert Koch 1878 seine Untersuchungen zur „Ätiologie der Wundinfektionen“ vorstellte, konnte er nicht ahnen, dass eines der von ihm beobachteten kokkoiden Elemente zu einem der wichtigsten Krankheitserreger zählt. Zur wesentlichen Entdeckung von Staphylokokken hat der schottische Chirurg Alexander Ogston beigetragen. Ogston berichtete in seinem berühmten Vortrag „Über Abscesse“ auf dem 9. Kongress der Deutschen Chirurgischen Gesellschaft in Berlin über eine Vielzahl von traubenförmig aussehenden Kokken in verschiedenen Eiterungsprozessen und gab diesen Mikroorganismen schließlich den Namen *Staphylococcus* [112]. Diese Bezeichnung leitet sich aus den griechischen Wörtern *staphyle* (Weintraube) und *kokkos* (Kugel) ab. 1884 klassifizierte der Chirurg F.J. Rosenbach die aus Eiter isolierten Staphylokokken aufgrund ihrer Pigmentierung auf Agarmedien in die zwei Arten, *Staphylococcus aureus* (lat. *aureus*=golden) und *Staphylococcus albus* (lat. *albus*=weiß) [52].

Bis aber zwischen pathogenen und apathogenen Staphylokokken unterschieden werden konnte, vergingen einige Jahrzehnte. Erst als J.v.Darányi 1926 einen Zusammenhang zwischen der Plasmakoagulase-Aktivität von Staphylokokken und der damit verbundenen pathogenen Bedeutung erkannte, wurde erstmals ein Schema zur Klassifizierung erstellt. Es sollte dennoch weitere 22 Jahre (1948) dauern, bis diese Beobachtungen auch allgemein akzeptiert wurden und *Staphylococcus aureus* als die humanpathogene Staphylokokkenart diagnostiziert werden konnte [52]. Plasmakoagulase ist ein Virulenzfaktor, mit dem - analog zum Thrombin - die Umwandlung von Fibrinogen zu Fibrin im Blutplasma unterschiedlicher Spezies katalysiert wird. Die Staphylokokkenspezies mit der höchsten pathogenen Potenz sowie der größten klinischen Bedeutung für den Menschen ist die koagulasepositive Spezies *S. aureus*.

Gegenwärtig beinhaltet das Genus *Staphylococcus* 72 weitere Spezies und Subspezies ([http://old.dsmz.de/microorganisms/bacterial\\_nomenclature\\_info.php?genus=Staphylococcus](http://old.dsmz.de/microorganisms/bacterial_nomenclature_info.php?genus=Staphylococcus)). Staphylokokken werden zur Familie der Staphylococcaceae aus der Ordnung der Bacillales gezählt, welche sich durch die fakultativ anaerobe Lebensweise

auszeichnet. Ausnahme ist *Staphylococcus aureus* subsp. *anaerobius*, eine an das Schaf adaptierte, obligat anaerobe Subspezies.

Die Differenzierung von *S. aureus* geht auf strukturelle Merkmale der Zellwand sowie physiologische, biochemische und molekulare Eigenschaften zurück. Staphylokokken gehören zu den Katalase-Bildnern und können so leicht von den Katalase-negativen Streptokokken abgegrenzt werden. Staphylokokken haben einen Gram-positiven Zellwandaufbau, sind unbeweglich und haben einen Durchmesser von ca. 0,5-1,5 µm.

Sie sind sehr widerstandsfähig und können unter verschiedenen Umweltbedingungen wachsen, wobei ihr Temperaturoptimum zwischen 30 °C und 37 °C liegt (Abb.1.1). Obwohl sie keine Sporen ausbilden, weisen sie dennoch eine weitgehende pH-Toleranz und Resistenz gegen Austrocknung auf und können kurzzeitig Temperaturen bis zu 60 °C überleben.



Abbildung 1.1: *Staphylococcus aureus*, ST121 Isolat nach einem Tag Wachstum (37 °C) auf Mueller-Hinton-Blutagar

### 1.1.2 Verbreitung und Bedeutung von *S. aureus* als Krankheitserreger

*S. aureus* ist weltweit verbreitet und gehört zu den gewöhnlichen Besiedlern von Haut, Hautdrüsen sowie den Schleimhäuten des Oropharynx bei Menschen, Säugetieren und Vögeln (Abb.1.2) [56, 146]. Als Besiedler der Haut wird *S. aureus* häufig im Perineum, Axilla oder in der Leistengegend gefunden und kann darüber hinaus auch als ein Teil der natürlichen Flora des Vaginal- und Verdauungstraktes isoliert werden [157, 139]. Die Fähigkeit zur Ausbildung von Biofilmen auf Polymeroberflächen und eukaryotischen Geweben stellt einen wichtigen Virulenzfaktor von *S. aureus* dar. Dadurch ist es *S. aureus* möglich, auch Kunststoff- und Metalloberflächen wie zum Beispiel Katheter, Implantate oder Shunts zu kolonisieren, die dann wieder als Infektionsquelle dienen können. Staphylokokken sind fakultativ pathogen, da eine Kolonisation zwar die Voraussetzung für eine Infektion ist und das Risiko dafür massiv steigert, aber nicht zwangsläufig dazu führen muss [72, 123]. Longitudinalstudien konnten zeigen, dass 20-75 % vorübergehend Träger, 10-35 % dauerhaft Träger sind und 5-50 % der Individuen nie von *S. aureus* besiedelt werden [73].

Staphylokokken gehören damit zu den wichtigsten Erregern von Infektionskrankheiten.

Gelingt es dem Keim, eine Eintrittspforte in den Organismus zu finden oder sich auf-

grund von Hygienemängeln oder Prädispositionen wie z.B. Immunsuppression, Diabeteserkrankung, bei Dialysepatienten oder in Folge einer HIV-Infektion auszubreiten, kann dies zu einer Reihe von Infektionen führen [88]. Diese Infektionen können in invasive (lokal, tiefgehend, systemisch) und toxin-vermittelte Erkrankungen eingeteilt werden [88]. Zu den invasiven Infektionen zählen u.a. Abszesse, Furunkel, Empyeme, Wund- und Weichteilinfektionen, Osteomyelitis, Pyodermie und Pneumonie. Ziehen oberflächliche und tiefgehende Prozesse eine Keimeinschwemmung in die Blutbahn nach sich, kann dies zu einer Endokarditis oder Sepsis führen [52]. Bei den toxinbedingten Erkrankungen lösen sezernierte Toxine eine direkte Reaktion des Wirtes aus. Eine typische toxin-vermittelte *S. aureus* Erkrankung ist die staphylogene toxische epidermale Nekrolyse (*Staphylococcal Scalded Skin Syndrom*, (SSSS)). Sie wird durch sezernierte Exfoliatine ausgelöst, die für eine intraepidermale Spaltbildung zwischen dem Stratum corneum und dem Stratum granulosum verantwortlich sind. Wenn eine Toxinausschüttung verhindert ist, treten oft nur deren lokalisierten Verlaufsformen *Pemphigus neonatorum* und *bullöse Impetigo* auf [52].

Des Weiteren können die von *S. aureus* sezernierten und mit der Nahrung aufgenommenen (Lebensmittelintoxikationen) Enterotoxine, die als Superantigene wirken, auch Symptome wie Übelkeit, Erbrechen und Durchfall herbeiführen. Auch das durch das Superantigen Toxische-Schock-Syndrom-Toxin-1 (TSST-1) ausgelöste Toxische-Schock-Syndrom (TSS), zählt mit seiner Fülle an klinischen Symptomen (Hautrötung, Fieber, Schock usw.) zur toxin-vermittelten Erkrankung durch *S. aureus* [88]. Der Verlauf und die Schwere der Erkrankung wird dabei durch eine massenhafte Ausschüttung von Zytokinen beeinflusst.

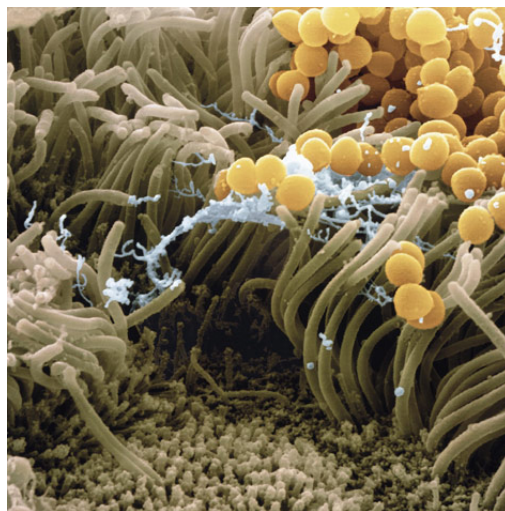


Abbildung 1.2: Rasterelektronenmikroskopaufnahme von *S. aureus* Bakterien (gelb) die humane Epithelzellen der Nase besiedeln. Die Anhaftung wird durch den Schleim (blau) der Zilien, die aus den Epithelzellen herausragen, gewährleistet. (Juergen Berger / Science Photo Library).

### 1.1.3 Pathogenität und Virulenzeigenschaften

Die Ursache für die Vielfalt der durch Staphylokokken hervorgerufenen pathogenen Prozesse und Infektionen liegt in der großen Anzahl von Virulenzfaktoren. Um bei der Besiedlung die Immunabwehr des Wirts zu umgehen, kann *S. aureus* auf ein breites Spektrum von vielen verschiedenen Pathogenitäts- und Virulenzfaktoren zurückgreifen. Diese Faktoren lassen sich in Zellwand-assoziierte und sezernierte Proteine einordnen. Durch Zellwand-assoziierte Proteine ist *S. aureus* in der Lage, an extrazelluläre Gewebematrix (Kollagen, Fibronectin, Laminin) zu binden und so eine

Kolonisation einzuleiten. Für weitere invasive Prozesse sind zytotoxische, membranzerstörende Proteine sowie Superantigene verantwortlich. In der Tabelle 1.1 ist eine Übersicht über die verschiedenen Faktoren dargestellt die von *S. aureus* zur Adhärenz, Invasion und Immunmodulation eingesetzt werden. Die Expression dieser Faktoren setzt eine komplexe Regulation intra- als auch extrazellulärer Signale voraus. Bei der Expression von Virulenzfaktoren greift *S. aureus* hauptsächlich auf das *agr* (*accessory gene regulator*)- System sowie das *sar* (*staphylococcal accessory regulator*)- System zurück [118, 109].

Verschiedene Versuche, bei denen das Kerngenom sowie das Repertoire an vorhandenen Virulenzgenen mit Isolaten von Kommensalen und Isolaten aus Blutkulturen verglichen wurden, scheiterten bei dem Versuch, eindeutige Determinanten für die zu erwartende Virulenz eines Stammes zu identifizieren [39, 60, 85]. Trotzdem sind mittlerweile einige Virulenzfaktoren bekannt die für spezifische Toxin-vermittelte Infektionen verantwortlich sind [41]. Aufgrund der häufigen Assoziation von Pantone-Valentine Leukozidin (PVL) und Exfoliativen Toxinen mit CC121-Isolaten konzentrieren sich die folgenden Abschnitte speziell auf diese beiden Virulenzfaktoren.

Tabelle 1.1: Pathogenitäts- und Virulenzfaktoren von *S. aureus* (in Anlehnung an [109])

Virulenzfaktor	Gen	Funktion
Clumping Factor A & B	<i>clfA, clfB</i>	Adhäsion (Fibrinogen)
Fibronectin-Bindungsproteine	<i>fna, fnb</i>	Adhäsion (Fibrinektin)
Kollagen-Bindungsprotein	<i>cna</i>	Adhäsion (Kollagen)
Elastin-Bindungsprotein	<i>ebpS</i>	Adhäsion (Elastin)
Fibrinogen-Bindungsprotein	<i>fbp</i>	Adhäsion (Fibrinogen)
Extrazelluläres	<i>efb</i>	Adhäsion (Fibrinogen)
Fibrinogen-Bindungsprotein		
Laminin-Bindungsprotein	<i>lbp</i>	Adhäsion (Laminin)
Koagulase	<i>coa</i>	Adhäsion (Fibrinogen, Prothrombin)
Intrazelluläre Adhärenz-Proteine	<i>icaA-D, icaR</i>	Adhäsion (Biofilmbildung)
Kapsel	<i>cap8A-cap8P</i>	Antiphagozytose (Kapselbildung)
Protein A	<i>spa</i>	Adhäsion/Antiphagozytose (Bindung der Fc-Domäne von IgG)
$\alpha$ -Hämolysin	<i>hla</i>	Membranzerstörung
$\beta$ -Hämolysin	<i>hlb</i>	Zerstörung von Blutkörperchen
$\gamma$ -Hämolysin	<i>hlgA-C</i>	Membranzerstörung
$\delta$ -Hämolysin	<i>hld</i>	Membranzerstörung
Leukozidine, PVL	<i>lukE-D,</i> <i>lukS/F-PV</i>	Zerstörung von polymorphkernigen Granulozyten und Makrophagen
Staphylokinase	<i>sak</i>	Zellinvasion
CHIPS	<i>chp</i>	Repression der Chemotaxis
SCIN	<i>scn</i>	inhibiert Komplementaktivierung
Hyaluronidase	<i>hysA</i>	Zellinvasion (löst Hyaluronsäure auf)
Glycerolester-Hydrolase	<i>geh</i>	Gewebezerstörung
Lipase	<i>lip</i>	Gewebezerstörung, Zellinvasion
Phospholipase C	<i>plc</i>	Gewebezerstörung, Zellinvasion
Proteasen	<i>spsA-B</i>	Gewebezerstörung

Thermonuklease	<i>nuc</i>	Zellinvasion (Degradation von Wirtsnukleinsäuren)
Enterotoxine	<i>sea-see, seg-seu</i>	Lebensmittelintoxikationen
Exfoliative Toxine A-D	<i>eta,etb,etc,etd</i>	Zerstörung von Zellmembranen
TSST-1	<i>tst</i>	induziert Toxisches-Schock-Syndrom

---

**Exfoliative Toxine.** Vorwiegend oberflächliche Infektionen wie Impetigo bullosa, exfoliative Dermatitis und SSSS sind häufig mit den Exfoliativen Toxinen A und B (ETA und ETB) assoziiert. Diese Hautinfektionen betreffen dabei ausschließlich die oberen Schichten der Haut innerhalb der Epidermis (*stratum corneum* und *stratum granulosum*) [14].

ETA und ETB zeigen eine strukturelle Verwandtschaft mit Serinproteasen und sind mit der Bindung und dem Abbau von desmosomalen Cadherinen assoziiert, insbesondere Desmoglein-1 (Dsg-1) [24, 149, 148, 5]. Desmoglein-1 und Desmoglein-3 gehören zusammen mit anderen desmosomalen Cadherinen zu den wichtigsten Haftproteinen in der Haut. Sie vermitteln in Epidermis und oberflächennahen Schleimhäuten über homo- und heterophile extrazelluläre Bindungen den Zell-Zell-Kontakt. Dsg-1 und Dsg-3 besitzen dabei ein umgekehrtes Expressionsmuster in der Epidermis. Während Dsg-1 in allen Hautschichten exprimiert wird, ist Dsg-3 nur in den tieferen Hautschichten exprimiert [122, 4]. Aufgrund der spezifischen Aufspaltung von Dsg-1 durch die Toxine ETA und ETB, kommt es zur Auflösung der Zell-Zell-Kontakte zwischen *Stratum spinosum* und *Stratum granulosum* innerhalb der Epidermis und dadurch zur Blasenbildung und Exfoliation der oberen Hautschichten [14]. Die unteren Hautschichten sind hierbei nicht betroffen, da Dsg-3 die Aufspaltung von Dsg-1 durch die Exfoliativen Toxine kompensiert [14]. Dieser Prozess ist typisch für das *Staphylococcal scalded skin syndrome*, dass auch Dermatitis exfoliativa Ritter von Rittershain genannt wird. Diese Erkrankung wurde erstmals 1878 von Baron Gottfried Ritter von Rittershain beschrieben und tritt meist bei Säuglingen und Kleinkindern auf [128].

Beide Toxine werden durch mobile genetische Elemente vermittelt. Während das Gen *eta* ausschließlich durch Prophagen übertragen wird, ist das Gen *etb* plasmidal kodiert [154, 42].

**Panton-Valentine Leukozidin.** Die Klasse der Leukozidine fasst mehrere porenbildende Toxine zusammen. Leukozidine helfen *S. aureus* die zelluläre Immunantwort zu umgehen und insbesondere polymorphkernige Leukozyten irreversibel zu schädigen [43]. Von besonderer Bedeutung ist das sogenannte Panton-Valentine Leukozidin (PVL), dass vor allem bei Infektionen durch caMRSA eine besondere Stellung einnimmt und oft als Charakteristika von caMRSA gesehen wird [117]. Erstmals wurde PVL 1894 von Van de Velde beschrieben, doch erst 1932 konnten

Panton und Valentine auch einen kausalen Zusammenhang zwischen der Ausbildung von Abszessen und der Produktion von PVL belegen [22, 119, 147].

PVL ist ein aus zwei Komponenten bestehendes porenbildendes Toxin. Die dafür kodierenden Gene *lukS-PV* und *lukF-PV* werden durch verschiedene Prophagen vermittelt, welche gleichzeitig auch zur Verbreitung des Toxins innerhalb der *S. aureus* Population beitragen [158]. Derzeit sind 10 PVL-kodierende Prophagen detailliert in der Literatur beschrieben, die alle der Virusfamilie *Siphoviridae* angehören:  $\Phi$ PVL,  $\Phi$ 108PVL,  $\Phi$ SLT,  $\Phi$ Sa2mw,  $\Phi$ Sa2958,  $\Phi$ Sa2usa,  $\Phi$ 7247PVL,  $\Phi$ 5967PVL,  $\Phi$ SLT-USA300\_TCH1516 und  $\Phi$ tp310-1 [158, 163].

Zunächst werden beide Determinanten sekretiert, bevor sie sich zu einem porenbildenden Heptamer assemblieren. Dieser führt anschließend zur Lyse von Leukozyten. Darüber hinaus kommt es zur Aktivierung zahlreicher anderer Mechanismen wie die vermehrte Freisetzung von chemotaktischen Faktoren, die zur Vasodilatation und Gewebnekrose führen, vermutlich ausgelöst durch freigesetzte reaktive Sauerstoffderivate oder Entzündungsmediatoren [64, 10, 89].

#### 1.1.4 Antibiotikaresistenzen

Mehr als 100 Jahre nachdem Alexander Ogston bereits die Bedeutung von Erkrankungen durch Staphylokokken in seinen Arbeiten erkannte und publizierte, gehört *S. aureus* zu einem der wandlungsfähigsten und bedeutendsten Infektionserreger im humanmedizinischen Bereich. Erst mit der Entdeckung des Antibiotikums Penicillin durch Alexander Fleming im Jahr 1928 und dessen Einführung im Jahr 1941 war die Therapie bakterieller Infektionen ohne eine chirurgische Versorgung möglich. Es wurde als „*Wundermedikament*“ und als einer der größten medizinischen Fortschritte gefeiert, da mit Penicillin sonst letal verlaufende Infektionen beherrschbar wurden. Doch bereits drei Jahre später konnte über die ersten auf Grund von  $\beta$ -Laktamase Bildung penicillinresistenten Staphylokokken berichtet werden [8]. In den darauffolgenden Jahren wurde eine Vielzahl von Antibiotika entwickelt, doch so schnell wie diese entwickelt und therapeutisch eingesetzt wurden, so rasant verlief auch die Evolution der dazugehörigen Resistenzen (siehe Abbildung 1.3). So konnte nur zwei Jahre nach der Einführung des Methicillins (halbsynthetisches,  $\beta$ -Lactamase-stabiles Penicillin) im Jahr 1961 bereits das Auftreten eines Methicillin-resistenten *S. aureus* (MRSA) in Großbritannien beobachtet werden [63]. Dies war die Geburtsstunde von MRSA, der sich fortan zu einem Problem in der Human- und Veterinärmedizin mit zunehmenden globalen Charakter entwickeln sollte. Neben dem Auftreten von MRSA in medizinischen Einrichtungen (sog. haMRSA, *hospital-associated*), werden diese auch unabhängig von medizinischen Einrichtungen beschrieben (sog. caMRSA, *community-associated*). In den letzten Jahren wird zusätzlich von MRSA berichtet, die vor allem mit der Tiermast assoziiert sind (sog. laMRSA, *livestock-associated*). Trotz der Entwicklung vieler neuer Antibiotika bleibt die Behandlung von Staphylo-

kokken-Infektionen nach wie vor schwierig. Der vermehrte Einsatz von Antibiotika und die damit verbundene Erhöhung des Selektionsdrucks hat die Evolution der Resistenzentwicklung beschleunigt und zur Ausbildung vielfältiger Resistenzen geführt (siehe Abb. 1.3) [19]. Diese Resistenzentwicklung wird zum einen durch Mutationen und zum anderen durch den horizontalen Gentransfer von verschiedenen mobilen genetischen Elementen (z.B. SCC<sub>mec</sub>-Kassette, Resistenzgenen, Plasmide usw.) sowohl interspezifisch als auch intraspezifisch vorangetrieben [87, 57]. Dadurch gelang es *S. aureus* immer wieder neue MRSA-Klone hervorzubringen, die parallel zur Methicillinresistenz noch zusätzliche Resistenzen wie z.B. gegen Aminoglykoside, Makrolide, Tetracycline, Lincosamide, Fluorochinolone, Sulfonamide und andere Substanzklassen aufweisen.

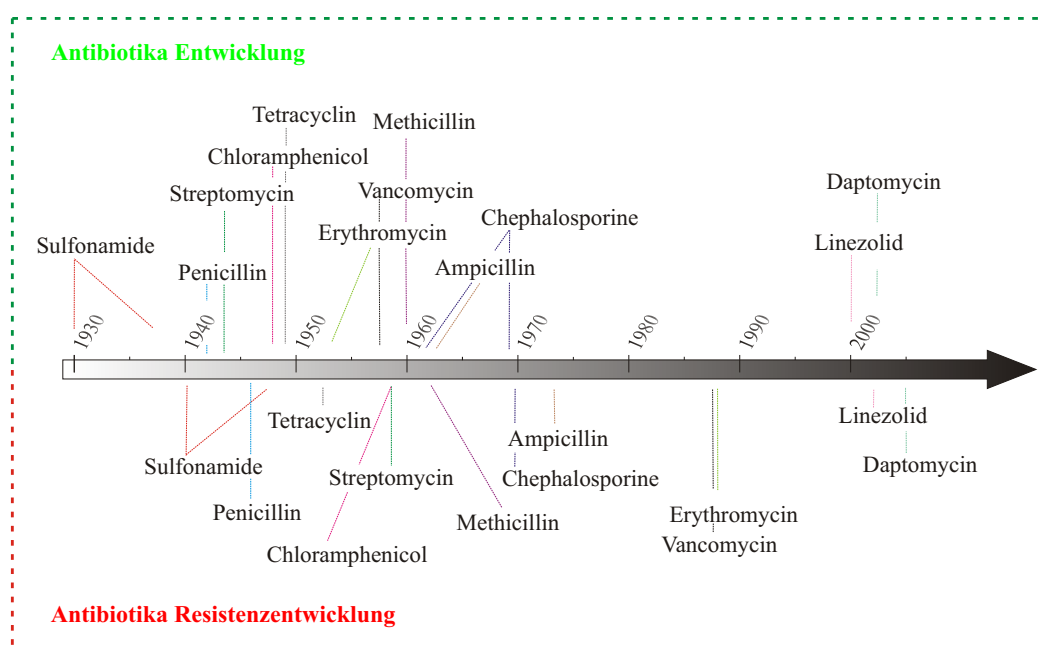


Abbildung 1.3: Zeitachse zur Antibiotikaentwicklung und der Ausbildung von Resistenzen. Oberhalb der Zeitachse ist die Antibiotikaentwicklung dargestellt, während unterhalb das Auftreten der ersten zugehörigen Resistenzen dargestellt wird (in Anlehnung an [19]).



### 1.1.5 Entwicklung von Methicillin-resistenten *Staphylococcus aureus*

Die Methicillinresistenz wird durch ein zusätzliches induzierbares, 76-kDa großes Penicillinbindepotein (PBP2a) verursacht, das vom Gen *mecA* kodiert wird [62]. Da PBP2a ein Protein mit niedriger Affinität zu  $\beta$ -Laktamen ist, bleibt die katalytische Funktion erhalten, selbst wenn die anderen vier membrangebundenen PBPs in *S. aureus* bereits durch ein  $\beta$ -Laktam Antibiotika inhibiert worden sind [125]. Dieser Resistenzmechanismus führt zum Wirkverlust aller  $\beta$ -Laktame, einschließlich Cephalosporinen, Carbapenemen und Monobactamen [67]. Bei der Entwicklung von einem Methicillin-sensitiven *S. aureus* (MSSA) zu einem Methicillin-resistenten *S. aureus* spielt das auf dem mobilen genetischen Element (MGE) SCC*mec* lokalisierte *mecA*-Gen, als eine zentrale genetische Determinante, eine entscheidende Rolle. Das SCC*mec*-Element kommt als genomische Inselstruktur in allen MRSA-Stämmen vor und inseriert immer an der gleichen Stelle im *S. aureus*-Genom, am 3'-Ende des *orfX*-Gens [62]. Das SCC*mec*-Element besteht neben dem *mecA* Gen noch aus weiteren Genen. Dabei handelt es sich unter anderem um  $\beta$ -Laktamase-Regulatoren, um die Rekombinase-Gene *ccrA*, *ccrB* und *ccrC* und um die Bereiche zwischen und neben den *mec*- bzw. *ccr*-Komplexen, die sogenannte J-Regionen [18, 62]. Die Rekombinase-Gene sind aus der Familie der Invertasen-Resolvasen und für den Transfer des SCC*mec*-Elementes notwendig, während die J-Regionen oftmals zusätzliche Resistenz-Gene enthalten [68, 69, 57].

Als Anfang der 60er die ersten MRSA-Stämme in Dänemark auftraten, wurde lange vermutet, dass die MRSA-Evolution dort ihren Ursprung fand und sich anschließend MRSA-Klone sehr schnell global ausgebreitet hatten [23]. Doch die Entdeckung unterschiedlicher SCC*mec*-Elemente in diversen klonalen Linien von *S. aureus* machte deutlich, dass die SCC*mec* Genkassetten von unterschiedlichen *S. aureus*-Linien aufgenommen und weiterverbreitet wurden [111, 81]. Derzeit können 11 verschiedene SCC*mec* Typen unterschieden werden, was das Potential für die Entwicklung von MRSA klar verdeutlicht (*International Working Group on the Staphylococcal Cassette Chromosome elements*, <http://www.sccmec.org/>).

Bis heute ist die Frage nach dem Ursprung des *mecA*-Gens nicht eindeutig geklärt. In einige Studien wird vermutet, dass der Ursprung des *mecA*-Gens in Koagulase-negativen Staphylokokken liegen könnte [160, 57].

Bei den Untersuchungen zur Prävalenz von MRSA gibt es sowohl national als auch international erhebliche Unterschiede. Abbildung 1.4 verdeutlicht diese innerhalb Europas. Während in nördlichen Ländern wie Niederlande, Island, Norwegen oder Schweden die Prävalenz <1 % liegt, steigt diese in südlichen Ländern wie Portugal, Frankreich, Italien und Spanien bis auf 20-50 % an. In den USA und Japan kann eine MRSA-Rate von 30-80 % nachgewiesen werden und auch Australien und mehrere Ländern in Afrika und Südamerika zeigen eine ähnlich hohe Prävalenz [15, 145].

Die am weitesten verbreitete klonale Linie in Deutschland gehört dem Sequenztyp 22 (Barnim-Epidemiestamm) an und ist dicht gefolgt vom Sequenztyp 225 (Rhein-Hessen-Epidemiestamm) [61].

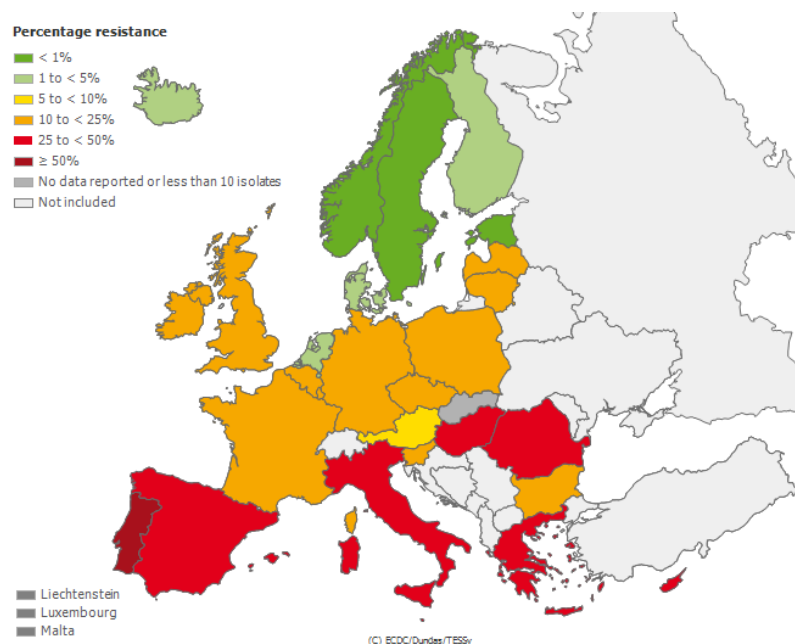


Abbildung 1.4: Anteil von MRSA an *S. aureus*-Isolaten aus Blutkulturen in europäischen Krankenhäusern im Jahr 2010. (Quelle: EARS-Net Datenbank ([www.ecdc.europa.eu](http://www.ecdc.europa.eu)))

### 1.1.6 Genomstruktur von *S. aureus*

Das Genom von *S. aureus* kann in das Kern- und akzessorische Genom unterteilt werden. Dabei umfasst das Kerngenom etwa 75 % der gesamten Erbinformation und enthält vorwiegend Gene für den zentralen Metabolismus [94, 84]. Darüber hinaus gehören aber auch Gene dazu, die Oberflächenproteine, Toxine und andere Exoenzyme kodieren die speziesspezifisch vorkommen [84].

Das akzessorische Genom, das keine essentiellen Aufgaben übernimmt, macht die restlichen 25 % des Genoms aus [84, 94]. Den Großteil des akzessorischen Genoms stellen mobile genetische Elemente (MGEs), die horizontal übertragen werden und so hauptsächlich zur Variabilität innerhalb der *S. aureus*-Population beitragen [84]. Zu diesen Elementen gehören Insertionssequenzen, Transposons, Prophagen, Pathogenitätsinseln, genomische Inseln, Plasmide und chromosomale Kassetten (SCC*mec*). Das akzessorische Genom kann Virulenz-assoziierte Faktoren oder Resistenzdeterminanten kodieren und so zu einem gesteigerten Pathogenitäts-, Virulenz- oder Resistenzverhalten führen [84]. Da in *S. aureus* vor allem Prophagen für den horizontalen Transfer von Genen verantwortlich sind und somit eine primäre Stellung in der evolutionären Entwicklung einnehmen, hilft die Kenntnis über deren Verbreitung bei Schlussfolgerungen über Evolution und ihrer Rolle in der Pathogenese [84, 97].

### 1.1.7 Molekulare Typisierungsmethoden

Die Populationsstruktur von MRSA wird von einer geringen Anzahl klonaler Linien dominiert. Um diese epidemiologisch relevanten Stämme und deren charakteristische Unterschiede von epidemiologisch irrelevanten Stämmen differenzieren zu können, sind Methoden zur frühzeitigen Erkennung, Überwachung und Typisierung von *S. aureus* erforderlich.

Zu den wichtigsten genotypischen Methoden, die zur Untersuchung von MRSA eingesetzt werden [129, 28] zählen die Pulsfeld-Gelelektrophorese (PFGE) [50], Multi-Locus-Sequenz-Typisierung (MLST) [33], *SCC<sub>mec</sub>*- [2] und *spa*-Typisierung [53].

**Pulsfeld-Gel-Elektrophorese (PFGE).** Die 1984 von Schwartz und Cantor entwickelte Pulsfeld-Gel-Elektrophorese ist eine Makrorestriktionsanalyse, bei der mit Hilfe von selten schneidenden Restriktionsenzymen (z.B. *SmaI*) das gesamte Bakteriengenom verdaut wird [138]. Es entstehenden Fragmente mit unterschiedlichen Größen (10-700 kbp), die anschließend in einem gepulsten elektrischen Feld aufgetrennt werden. Durch das so entstehende charakteristische Bandenmuster, den Restriktionsfragmentpolymorphismus (RFLP), können Bakterienstämme unterschieden werden.

**Multilocus-Sequenz-Typisierung (MLST).** Die Entwicklung von MLST begann mit der Untersuchung von *Neisseria meningitidis* und entwickelte sich schnell auch zu einem exzellenten Werkzeug für die Erforschung der klonalen Evolution von MRSA [91, 33]. MLST basiert auf der Sequenzanalyse von PCR-Fragmenten von sieben *S. aureus* Haushaltsgenen. Diese umfassen dabei die Carboxylkinase *arcC*, die Shikimat-Dehydrogenase *aroE*, die Glycerolkinase *glpF*, die Guanylatkinase *gmk*, die Phosphat-acetyltransferase *pta*, die Triosephosphatisomerase *tpi* und die Acetylcoenzym A Acetyltransferase *yqiL* [33]. Durch diese Analysen ergeben sich für verschiedene Isolate unterschiedliche Sequenzen in den entsprechenden Allelen jedes einzelnen Haushaltsgens, so dass als Resultat ein spezifisches Allelprofil bzw. ein charakteristischer Sequenztyp (ST) für jedes Isolat definiert werden kann [29].

***Staphylococcus-aureus*-Protein-A-Typisierung (*spa*-Typisierung).** *S. aureus* bildet das durch das *spa*-Gen kodierte Oberflächenprotein Protein A aus. Innerhalb des *spa*-Gens befindet sich eine hypervariable X-Region, die aus DNA *Repeats* aufgebaut ist. Diese *Repeats* sind 24 Bp lang und können in ihrer Anzahl und Sequenz im Chromosom von Stamm zu Stamm variieren und sind so für die Differenzierung geeignet [141, 53].

### 1.1.8 Klonale Populationsstruktur

Das zunehmende Auftreten von MRSA und die daraus resultierende Morbidität, Mortalität sowie die damit verbundene finanzielle Belastung im Gesundheitswesen haben ein signifikantes Interesse an populationsbiologischen Fragestellungen bezogen auf MRSA aufkommen lassen [155, 111]. Für die USA konnten im Zeitraum von 1999-2005 retrospektiv über 450.000 *S. aureus*-bedingte Hospitalisationen mit circa 10.800 Todesfällen jährlich beobachtet werden [71]. Allein 2003 verursachten alle stationär behandelten *S. aureus*-Infektionen eine finanzielle Belastung von über 14 Milliarden Dollar [105].

Durch den Einsatz von Typisierverfahren wie der Multilocus-Sequenz-Typisierung war es möglich, Populationsanalysen durchzuführen. Diese konnten zeigen, dass die Populationsstruktur von *S. aureus* stark klonal aufgebaut ist und die meiste Diversität der Population in einer limitierten Anzahl von klonalen Komplexen vorliegt (siehe Abbildung 1.5) [34].

Durch verschiedene Studien konnte gezeigt werden, dass die Vielzahl der heutigen Krankenhaus assoziierten *S. aureus*-Infektionen auf fünf klonale Komplexe (CC5, CC8, CC30, CC22 und CC45) zurückgeführt werden können, die sowohl endemisch als auch pandemisch auftreten [114, 48, 34]. Nachdem dieses Phänomen auch in anderen Spezies beobachtet werden konnte, stellte sich zunehmend die Frage, was diese klonalen Komplexe so erfolgreich macht. Einige klonale *S. aureus* Linien (z.B. CC22, ST239) erlangten durch die Akquirierung einer SCC*mec*-Kassette einen Selektionsvorteil gegenüber anderen und konnten sich so erfolgreich global ausbreiten [23, 57, 55]. Mit der Ent-

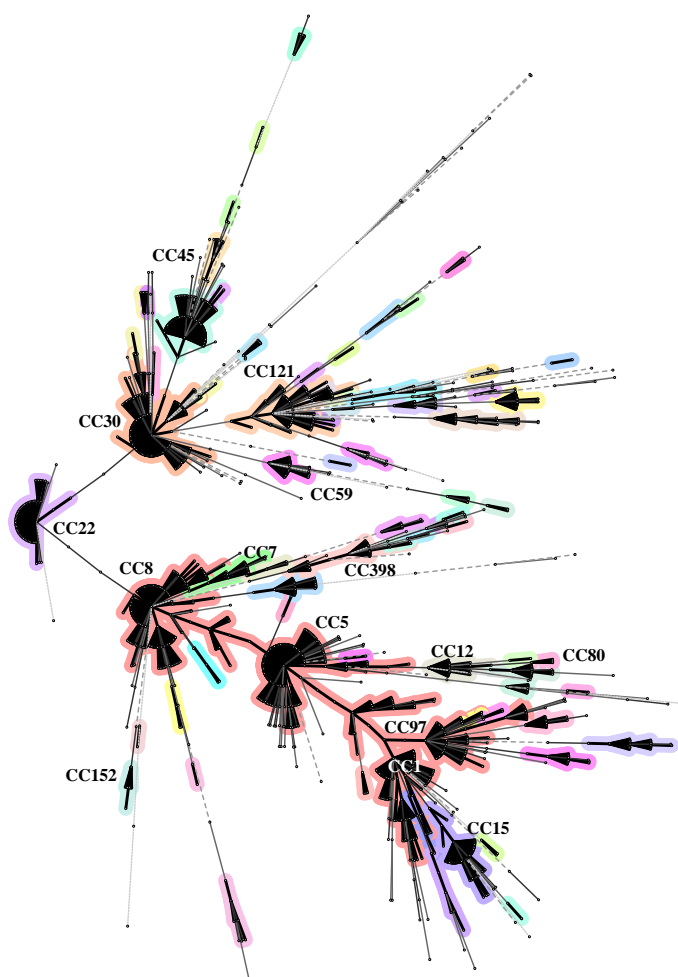


Abbildung 1.5: Populationsstruktur von *S. aureus* basierend auf MLST-Daten (Quelle: persönliche Mitteilung von B. Strommenger 2012).

deckung verschiedener SCC*mec*-Typen in gleichen Genotypenlinien und den Erkenntnissen anderer Studien, konnte ein anderes Bild über die molekulare Evolution

von MRSA gezeichnet werden. So konnte gezeigt werden, dass eine Rekombination zwischen SCC*mec*-Elementen ein eher seltenes Ereignis ist und die verschiedenen Varianten weitestgehend stabil erhalten bleiben [81, 18]. Demnach trifft die Auffassung pandemischer MRSA-Klone nicht für alle epidemischen MRSA-Klone zu. So haben verschiedene global auftretende MSSA-Linien unabhängig voneinander lokal unterschiedliche SCC*mec*-Elemente erworben, aus denen sich anschließend mehrere MRSA-Linien entwickeln konnten die sich nicht weiter ausbreiten [48, 129, 111]. Dennoch bleibt die Aufnahme Frequenz von SCC*mec* und anderen Virulenzfaktoren sowie den vielen Faktoren, die für die Dominanz besonderer Klone sorgen, bis heute ungeklärt.

Aufgrund der limitierten Diskriminierungsfähigkeit der derzeit für *S. aureus* verwendeten Typisierverfahren ist es nicht möglich, tiefergehende Fragen zur Evolution von *S. aureus* zu beantworten. Im Gegensatz zur Spezies *Neisseria meningitidis* sind auch keine Korrelationen von bestimmten klonalen Komplexen und spezifischen klinischen Phänotypen abzuleiten [144, 37]. Ein durch MLST festgelegter individueller Sequenz Typ (ST) kann in seinem Resistenzverhalten, *spa* Typ, SCC*mec*-Typ sowie in seinem Gengehalt durchaus variieren [83, 141, 159].

Weitaus detailliertere Analysen bieten dabei die Analysen von Einzelnukleotidpolymorphismen (engl. *single nucleotide polymorphism*, SNP) und Genom-weiten SNPs. Anhand dieser Analysen ist es möglich, sehr präzise Aussagen über die Evolution als auch demographische Entwicklungen in der *S. aureus*-Population zu rekonstruieren. Die Erforschung der Phylogenie anhand von SNP-Analysen wurde bereits für einige Krankheitserreger wie z.B. *Yersinia pestis* [1], *Mycobacterium tuberculosis* [3, 7], *Bacillus anthracis* [124] und *Salmonella* Typhi [131] aber auch für einige klonale *S. aureus*-Linien durchgeführt [111, 110, 55].

## 1.2 Sequenztyp 121, ST121

*S. aureus* Isolate, die dem klonalen Komplex CC121 zugeordnet werden können, gehören zu einer prädominanten klonalen Linie, die besonders in der Bevölkerung außerhalb von Krankenhäusern und anderen medizinischen Pflegeeinrichtungen als Hauptverursacher von Haut- und Weichteilgewebe-Infektionen (engl. *skin and soft tissue infections*, SSTIs) zählt [21, 126, 151, 77]. CC121 setzt sich aus den Sequenztypen ST121, ST123 und einer Reihe weiterer Varianten zusammen. Der Sequenztyp ST121 stellt die Ursprungspopulation von CC121 dar, nach dem der klonale Komplex auch benannt wurde. Dieser definiert sich durch das folgende MLST-Allelprofil: 3-6-5-6-2-7-14-5.

Internationale Studien aus Europa, Russland, Asien, Ozeanien, USA und Afrika belegen eine weltweite Verbreitung des ST121 Klons, sowohl bei asymptomatischen Trägern als auch in Infektionen [21, 126, 151, 77, 16, 100, 126, 96, 25].

*S. aureus*-Infektionen die mit CC121-Isolaten assoziiert sind, können aus klinisch morphologischer sowie therapeutischer Sicht in oberflächliche (Epidermis betreffend) und tiefgehende Infektionen (Dermis und Subkutis) unterschieden werden [93, 133, 116, 77, 126]. Auffällig dabei ist, dass diese zwei klinischen Krankheitsbilder häufig mit der Expression bestimmter Toxine assoziiert sind. So konnten oberflächliche Infektionen wie bullöse Impetigo und SSSS allgemein mit der Expression von Exfoliativen Toxinen (ETA und ETB) in Verbindung gebracht werden [14, 82, 49, 133], während tiefgehende Infektionen wie rezidivierende Abszesse und nekrotisierende Pneumonien vorwiegend mit dem Vorhandensein von PVL assoziiert sind [59, 82, 93]. In dem Kapitel 1.1.3 sind diese beiden charakteristischen Pathogenitätsfaktoren und ihre spezifischen Wirkmechanismen näher erläutert. In einer Studie aus Polen konnten in einem Zeitraum von 2002-2008, 55 % aller Furunkulose-Fälle, die durch MSSA-Isolate verursacht wurden, dem klonalen Komplex CC121 zugeordnet werden [93], was die enorme Bedeutung dieses Klons besonders außerhalb von Krankenhäusern und anderen medizinischen Pflegeeinrichtungen unterstreicht.

Ein weiterer sehr interessanter Aspekt ist, dass bis 2009 angenommen wurde, dass es in dieser *S. aureus*-Linie nie zu der Integration einer SCC<sub>mec</sub>-Kassette gekommen ist und alle Isolate damit generell empfindlich gegenüber dem Antibiotika Methicillin sind [28, 34, 59]. In den letzten drei Jahren wurden allerdings Methicillin-resistente CC121 Isolate aus dem asiatischem Raum beschrieben [16, 103, 38, 70].

## 1.3 Zielsetzung

Der Fokus der vorliegenden Arbeit liegt auf folgenden Schwerpunkten:

- Erforschung und Analyse der evolutionären Populationsstruktur des Klonalen Komplexes CC121 insbesondere des Sequenztyps ST121
- Untersuchung zur Fragestellung, ob bestimmte Subpopulationen innerhalb der ST121 Population mit bestimmten klinischen Phänotypen assoziiert sind
- Vergleichende Genomanalysen zur Aufklärung der genetischen Ausstattung und zur Überprüfung phylogenetischer Zusammenhänge
- Untersuchung von möglichen genomischen Mechanismen, die zur späten MRSA Evolution in CC121 beigetragen haben könnten

## Teil II

### Material und Methoden



## 2. Material

### 2.1 *Staphylococcus aureus* Isolate

Die 154 nachfolgend beschriebenen *S. aureus*-Isolate des klonalen Komplexes 121 stammen aus der Stammsammlung des Nationalen Referenzzentrums für Staphylokokken (NRZ) des Robert Koch-Instituts, oder wurden durch ausländische Kooperationspartner zur Verfügung gestellt. So konnten neben Isolaten aus Deutschland auch internationale *S. aureus* Isolate aus 27 verschiedenen Ländern und fünf Kontinenten untersucht werden. Dies war Voraussetzung, um eine möglichst große Diversität zwischen den Isolaten zu gewährleisten und um aussagekräftige Populationsanalysen durchführen zu können. Detaillierte Angaben zu den Stämmen befinden sich im Anhang in der Tabelle A.1.

## 2.2 Geräte

Tabelle 2.1: Geräte und Hersteller

Verwendete Geräte	Hersteller
Analysewaage	Sartorius
Autoklaven	Biomedis
Brutschränke, B 6760	Heraeus Instruments
Brutschränke, Innova <sup>®</sup> 42/42R	New Brunswick Scientific, Inc
dHPLC WAVE <sup>TM</sup> - System	Transgenomic Inc.
Elektrophoresekammer, Sub-Cell GT	BioRad
Elektrophoresekammer, Power Pac 300	BioRad
Geldokumentation, GelDoc XR	BioRad
Gefrierschränke -20 °C, LIEBHERRcomfort	Liebherr-holding GmbH
Kapillarsequenzierer, 3130xl Genetic Analyzer	Applied Biosystems, Darmstadt
Kühlschrank + 4 °C, Liebherr Comfort	Liebherr
Mikrowelle, MW81W	Samsung
Pipetten, Reference	Eppendorf AG
Pipettenspitzen (0,1 µl-5,0 µl)	Eppendorf AG
Pyrosequenzierer, 454 GS FLX	Roche, Mannheim
Sterilbank, HERAsafe HSP 9	Heraeus Instruments
Spannungsquelle, PowerPack Basic	Biorad
Thermocycler, GeneAmp PCR Systemm 9700	Applied Biosystems
Thermocycler, PTC-200 DNA Engine	Bio-Rad
Thermocycler, Labcycler	Sensoquest
Thermomixer, Comfort	Eppendorf AG
UV-Vis Spektrophotometer, BioPhotometer plus	Eppendorf AG
Vacuum Zentrifuge, Concentrator 5301	Eppendorf AG
Zentrifuge, Centrifuge 5804R	Eppendorf AG
Zentrifuge, Centrifuge 5417R	Eppendorf AG
Zentrifuge, MiniSpin <sup>®</sup> plus	Eppendorf AG

## 2.3 Chemikalien und Enzyme

Alle in der Tabelle 2.2 verwendeten Chemikalien sowie Enzyme wurden über die aufgeführten Firmen bezogen.

Tabelle 2.2: Chemikalien, Enzyme und Lieferanten

Chemikalien	Hersteller
Agarose	Sigma-Aldrich GmbH
Big Dye <sup>®</sup> Terminator v1.1 Cycle Seq. Kit	Applied Biosystems <sup>®</sup>
Bromphenolblau	Sigma-Aldrich GmbH
Destilliertes Wasser	Sigma-Aldrich GmbH
Ethanol (98 %)	Merck KGaA
Ethidumbromid (10 mg/ml)	Carl Roth GmbH + Co. KG.
EDTA (Ethyldiamintetraessigsäure)	Sigma-Aldrich GmbH
HCl (37 %)	Carl Roth GmbH + Co. KG.
Magnesiumchlorid MgCl <sub>2</sub>	Carl Roth GmbH + Co. KG.
NaCl (Natriumchlorid)	Carl Roth GmbH + Co. KG.
NaOH (Natriumhydroxid)	Carl Roth GmbH + Co. KG.

WAVE Puffer A	Transgenomic, Inc.
WAVE Puffer B	Transgenomic, Inc.
WAVE Wash Solution	Transgenomic, Inc.
WAVE Puffer D	Transgenomic, Inc.
Tris	Carl Roth GmbH + Co. KG.
TrisHCl	Merck KGaA
Triton X-100	Sigma-Aldrich GmbH

#### **DNA-Längenstandard**

GeneRuler 100 bp DNA-Ladder Plus	Fermentas GmbH
----------------------------------	----------------

#### **Enzyme**

Lysostaphin	Sigma-Aldrich GmbH
Optimase Polymerase	Transgenomic, Inc.
Proteinase K	Qiagen GmbH
RNaseI	Sigma-Aldrich GmbH
Taq-Polymerase	Invitrogen, Karlsruhe

## 2.4 Nährmedien

Folgende Nährmedien wurden verwendet.

Tabelle 2.3: Verwendete Nährmedien und Hersteller

Verwendete Nährmedien	Zusammensetzung	Hersteller
Luria Bertani +1 % Glycin (LB) Flüssigmedium	Trypton 10 g/l, Hefeextrakt 5 g/l, NaCl 0,5 g/l, Glycin (C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> ) 1 %, pH 7,0	Robert Koch-Institut
Mueller-Hinton (MH) Schafblutagar	Trypton 2 %, Hefeextrakt 0,5 %, NaCl 10 mM, KCl 2,5 mM, MgSO <sub>4</sub> 10 mM, Glukose 20 mM, MgCl <sub>2</sub> 10 mM, mit 5 % defibriniertem Schafblut	Oxoid GmbH
Tryptic Soy Broth (TSB)	Trypton 1 %, Hefeextrakt 0,5 %, NaCl 0,5 %, Glycin 1 % (C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> ), MgCl <sub>2</sub> 10 mM, Glycin (C <sub>2</sub> H <sub>5</sub> NO <sub>2</sub> ) 1 %, pH 7,4	BD Diagnostic Systems

## 2.5 Kits

Die verwendeten Kits sind in der Tabelle 2.4 aufgeführt.

Tabelle 2.4: Verwendete Kits

Verwendung	Kit-Name	Hersteller
DNA Isolierung	DNAeasy Blood and Tissue kit	Qiagen GmbH
Aufreinigung der PCR-Produkte	QIAquick PCR Purifikation Kit	Qiagen GmbH
Sequenzierung der PCR-Produkte	Big Dye <sup>®</sup> Terminator v1.1 Cycle Seq. Kit	GE Healthcare

## 2.6 Primer

Alle in der Arbeit verwendeten Primer zur Amplifizierung der einzelnen Gen-Loci oder zum Nachweis spezifischer Virulenzgene sind in den Tabellen A.4 und A.3 im Anhang aufgelistet und im Text entsprechend vermerkt.

## 2.7 Software und Internet-Tools

Tabelle 2.5: Verwendete Software-Pakete und Bezugsquellen

Software-Pakete	Verwendung	Bezugsquelle
Artemis Release 13.0	Annotation von Sequenzdaten	<a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a>
BaTs v1.3	Korrelationsanalysen von Phänotyp und Genotyp	<a href="http://evolve.zoo.ox.ac.uk">http://evolve.zoo.ox.ac.uk</a>
BEAST v1.6.1	Phylogenetische Analysen	<a href="http://beast.bio.ed.ac.uk">http://beast.bio.ed.ac.uk</a>
BEAUTi v1.6.1	Generieren von XML Dateien als Input für BEAST	<a href="http://beast.bio.ed.ac.uk">http://beast.bio.ed.ac.uk</a>
BioNumerics v6.5	Aufbau von Datenbanken	Applied Maths, Sint-Martens-Latem, Belgien
CorelDRAW v.12.0.0.536	Werkzeugassistent für die Bildbearbeitung	(c) 2003 Corel Corporation
DnaSP v5.10.01	Detektierung von homoplastischen SNPs	<a href="http://www.ub.edu/dnasp">http://www.ub.edu/dnasp</a>
FastaTOFastq.pl	Perl-Skript zum erstellen von FASTQ-Dateien	<a href="http://seqanswers.com/forums">http://seqanswers.com/forums</a>
FigTree v.1.3.1	Graphische Bearbeitung phylogenetischer Bäume	<a href="http://tree.bio.ed.ac.uk">http://tree.bio.ed.ac.uk</a>
Image Reader V1.5E	Dokumentation der Agarosegele	Fuji Foto Film
Kodon v3.6.1	Genom-Alignments; SNP Analysen	Applied Maths, Sint-Martens-Latem, Belgien
MEGA v5	Sequenzalignments	<a href="http://www.megasoftware.net">http://www.megasoftware.net</a>
MrBayes v3.1.2	Bayes'sche Analysen	<a href="http://mrbayes.sourceforge.net">http://mrbayes.sourceforge.net</a>
Navigator <sup>TM</sup>	Auswertung der dHPLC Ergebnisse	Transgenomic Inc., Lizenz
Software v3.0.0		Robert Koch-Institut
Path-O-Gen v1.3	Detektion des zeitlichen Signals in Sequenzdaten	<a href="http://tree.bio.ed.ac.uk">http://tree.bio.ed.ac.uk</a>
PAUP* v4.0b10	Berechnung des Homoplasie-Indexes (HI)	Sinauer Associates, Sunderland, USA (Wilgenbusch & Swofford 2003)
SAMtools v0.1.12a	Alignmentverarbeitung	<a href="http://samtools.sourceforge.net">http://samtools.sourceforge.net</a>
SeaView v4.2.3	ML-Methode und Alignment Tool	<a href="http://pbil.univ-lyon1.fr">http://pbil.univ-lyon1.fr</a>
SeqMan Pro v.8.1.5(3),418	Assemblierung von Sequenzdaten	DNASTAR Lasergene 8, Lizenz
SMALT v0.6.1	<i>Read Mapping</i> und Alignment Tool	Robert Koch-Institut
SSAHA2 v2.5.3	<i>Read Mapping</i>	<a href="http://www.sanger.ac.uk">http://www.sanger.ac.uk</a>
Texmaker 2.1	Erstellung von L <sup>A</sup> T <sub>E</sub> X-Dokumenten; Dissertation	<a href="http://www.xmlmath.net">http://www.xmlmath.net</a>
Tracer v1.5	Visualisierung des Beast-Outputs	<a href="http://beast.bio.ed.ac.uk">http://beast.bio.ed.ac.uk</a>

Tabelle 2.6: Verwendete Internet-Tools

<b>Internet-Tools</b>	<b>Verwendung</b>	<b>Bezugsquelle</b>
FaBox	Bearbeitung von FASTA-Dateien	<a href="http://users-birc.au.dk/biopv/php/fabox">http://users-birc.au.dk/biopv/php/fabox</a>
NCBI BLAST	Annotation sowie Multiple und paarweise Sequenzvergleiche	<a href="http://blast.ncbi.nlm.nih.gov">http://blast.ncbi.nlm.nih.gov</a>
PHAST	Annotation von Phagensequenzen	<a href="http://phast.wishartlab.com">http://phast.wishartlab.com</a>
Primer3 v0.4.0	Entwurf von Oligonukleotid Primern	<a href="http://frodo.wi.mit.edu">http://frodo.wi.mit.edu</a>
Pubmed	Literaturrecherche	<a href="http://www.ncbi.nlm.nih.gov/pubmed">http://www.ncbi.nlm.nih.gov/pubmed</a>
Reverse Complementary	Konvertierung von DNA in Komplementär-Sequenz	<a href="http://www.bioinformatics.org">http://www.bioinformatics.org</a>

## 3. Mikrobiologische Methoden

### 3.1 Bakterienanzucht und Stammhaltung

Die in der Arbeit verwendeten *S. aureus* Isolate wurden auf Mueller-Hinton-Blutagarplatten (Oxoid, Cambridge, UK) ausgestrichen, bei 37°C über Nacht kultiviert und für weitere Versuche bei 4°C aufbewahrt. Des Weiteren wurden alle Isolate in Form von Kryoproben asserviert und bei -80°C in der Stammsammlung des Robert Koch-Instituts aufbewahrt.

Einzelkolonien der Staphylokokken wurden für die Präparation der chromosomalen DNA aerob in Luria Bertani (LB) Flüssigmedium mit 1 % Glycin (LB-Medium+1 % Glycin, pH 7,4, je 4ml) sowie im späterem Verlauf der Arbeit in TSB-Flüssigmedium gegeben und bei 37°C, 150 rpm schüttelnd über Nacht inkubiert. Die Inkubation erfolgte in Reagenzgläsern mit 4ml LB-Flüssigmedium. So wurde ein konstantes Verhältnis von Kulturvolumen zum Gefäßvolumen von 1:5 gewährleistet. Anschließend wurde 1 ml der Übernachtskultur für die Isolierung der chromosomalen DNA weiterverwendet.

#### **Antibiotikaresistenzbestimmung**

Die minimalen Hemmkonzentrationen (MHK) verschiedener Antibiotika wurden mittels Mikrobouillonverdünnungstest nach DIN 58940 bestimmt. Der MHK-Wert entspricht der niedrigsten Konzentration einer antibiotischen Substanz, die das Wachstum des entsprechenden Keimes unter Versuchsbedingungen hemmt. Bei dem Test handelt sich um einen Reihenverdünnungstest, bei dem das Ergebnis anschließend visuell oder photometrisch beurteilt wird. Die ermittelten Resistenzen gegen Antibiotika sind für alle CC121-Isolate in der Tabelle A.1 aufgeführt. Diese phänotypische Resistenzbestimmung wurde von technischen Mitarbeiterinnen des Robert Koch-Institutes im Zuge der routinemäßigen Aufgaben des NRZ für Staphylokokken durchgeführt.

## 3.2 Präparation chromosomaler DNA

Die DNA wurde mittels Qiagen DNeasy Tissue Kit isoliert. Zur Extraktion der chromosomalen DNA von *S. aureus* wurde das entsprechende Extraktionsprotokoll des Kits verwendet. Abweichungen zum Protokoll bestanden in der Verwendung von 100 µg/ml Lysostaphin und nicht wie vorgesehen 20 mg/ml Lysozym. 1 ml einer Übernacht-Kultur wurden für 5 min bei 8000 rpm zentrifugiert und der Überstand verworfen. Das gewonnene Pellet wurde anschließend in 200 µl Lysispuffer (bestehend aus 2 mM EDTA, 20 mM Tris-HCL, 1,2 % Triton-X-100, 100 µg/ml Lysostaphin und pH 8,0) resuspendiert und für 30 min bei 37°C inkubiert. Im darauffolgenden Schritt erfolgte die Zugabe von 25 µl Proteinase K-Lösung (20 mg/ml) und 200 µl AL-Puffer. Nach gründlichem Mischen musste der Ansatz weitere 90 min bei 56°C inkubiert werden, um eine Spaltung der in dem Lysat befindlichen Polypeptide zu ermöglichen. Die Reaktion wurde anschließend mit der Zugabe von 200 µl 96 % Ethanol gestoppt, auf eine DNeasy Spin Column pipettiert und für 1 min in einer Tischzentrifuge bei 8000 rpm abzentrifugiert. Der Durchfluss wurde verworfen. Danach wurde auf die Säule 500 µl AW1-Puffer (Quiagen DNeasy Tissue Kit) pipettiert und in einem weiteren Zentrifugationsschritt für 1 min bei 8000 rpm abzentrifugiert. Ein weiteres Mal wurde der Durchfluss verworfen. Anschließend wurde auf die Säule 500 µl AW2-Puffer (Quiagen DNeasy Tissue Kit) pipettiert und für 3 min bei 14.000 rpm erneut abzentrifugiert. Zur Elution der DNA wurde die Säule für 2 min mit 100 µl A. bidest. inkubiert und anschließend für 1 min bei 8000 rpm zentrifugiert. Die Lagerung der isolierten DNA erfolgte für spätere Analysen bei -20 °C.

### 3.2.1 Bestimmung der DNA Konzentration

Die Konzentration der DNA-Präparationen wurde spektrophotometrisch im Bio-Photometer plus (Eppendorf AG) durch Absorptionsmessung bei 260 nm bestimmt. Eine Konzentration von 50 µg/ml doppelsträngiger DNA entspricht einer  $OD_{260} = 1$ . Die Reinheit der DNA-Präparation konnte rechnerisch anhand des Verhältnisses der Absorption bei 260 nm und 280 nm ( $A_{260}/A_{280}$ ) ermittelt werden. Der Quotient einer reinen DNA-Probe liegt bei 1,8; geringere Werte weisen auf eine Verunreinigung mit Protein, höhere auf eine Verunreinigung mit RNA hin.

### 3.3 Polymerase-Kettenreaktion (engl. *polymerase chain reaction*, PCR)

Die PCR wurde 1984 entwickelt und revolutionierte die Gentechnik [98]. Mittels der PCR ist eine Amplifizierung spezifischer DNA-Fragmente aller zu untersuchenden Isolate möglich.

#### PCR zur Detektion Virulenz-assoziiierter Toxingene (*eta*, *etb* und *lukS/F-PV*)

Zum Nachweis der Virulenzgene *eta*, *etb* und *lukS/F-PV*, wurde eine PCR verwendet. Die Exfoliativen Toxine A und B werden durch die Gene *eta* und *etb* kodiert, während die Gene *lukS-PV* und *lukF-PV* die Komponenten des Pantone-Valentine Leukozidin (PVL) kodieren. Die PCR-Reaktion wurde mittels PuReTaq Ready-To-Go PCR *Beads* (GE Healthcare Europe GmbH) realisiert. In den *Beads* liegen die Taq-Polymerase, die dNTPs und das zugehörige Puffersystem lyophilisiert vor. Lediglich die Zugabe von 24,25 µl A. bidest., 0,25 µl reaktionsspezifischer Primer F (10 µM), 0,25 µl reaktionsspezifischer Primer R (10 µM) und 0,25 µl Template auf ein Endvolumen von 25 µl ist notwendig. Die Amplifikationen erfolgten im Thermocycler PTC-200 DNA Engine (Bio-Rad) mit dem folgendem Programm:

- Anfangs Denaturierung: 96 °C, 5 min
  - Denaturierung: 96 °C, 30 s
  - Annealing: 55 °C, 45 s
  - Elongation: 72 °C, 60 s
  - Abschluss Elongation: 72 °C, 5 min
  - Kühlung: 8 °C
- } 30 Zyklen jeweils

Für die Überprüfung der Amplifikationsreaktion und zum Nachweis der jeweiligen Toxin-Gene (*eta*, *etb* und *lukS/F-PV*) wurden alle PCR-Produkte in einem 1.5 % Agarosegel mit 1 x TBE als Elektrophoresepuffer (40 mM Tris, 40 mM Borsäure, 1 mM EDTA, pH 8.3) aufgetrennt, mit Ethidiumbromid gefärbt und anschließend analysiert. Die Tabelle A.4 im Anhang enthält die verwendeten Oligonukleotid-Primer.

#### PCR zum Nachweis der spezifischen *attachment site* von Prophagen der Integrase-Gruppe Sa2

Für den Nachweis der spezifischen *attachment site* (*attB*; ACCATCTCATTATGATGATATGTTTATTT), die sich in einem Gen unbekannter Funktion (ORF-Bezeichnung in N315: SA1320) befindet und für die Integration von Prophagen der



Integrase-Gruppe Sa2 Voraussetzung ist [65], wurde eine Nachweis-PCR konzipiert. Mit Hilfe dieser Nachweis-PCR wurden alle 43 *S. aureus*-Isolate untersucht, in denen kein PVL nachgewiesen werden konnte (siehe Tab.A.1). Die komplementären Sequenzen für die Primerpaare lagen um die spezifische *attachment site* im codierenden Bereich des Gens SA1320. Nach Amplifikation betrug die PCR-Produktlänge 180 bp. Das entsprechende Primerpaar (*attB\_PVLint-L* und *attB\_PVLint-R*) wurde mit Hilfe des Programms Primer3 am Referenzstamm 09-00227 (ST121-Isolat) abgeleitet. Die Tabelle A.4 im Anhang enthält die verwendeten Oligonukleotid-Primer. Als optimale Bindungstemperatur wurde 55°C ermittelt, bei einer Elongationszeit von 1 min.

### PCR für die SNP-Analyse mittels dHPLC

Für die SNP-Analyse mittels der dHPLC wurde eine speziell dafür ausgerichtete Polymerase (Optimase-Polymerase) mit 3'→5'-Exonuklease-Aktivität verwendet. Die DNA-Präparationen wurden in einer Konzentration von 15 ng/µl eingesetzt. Bei den DNA-Amplifikationen wurden pro 25 µl Reaktionsansatz: 2,5 µl 10 x Reaktionsspuffer (PeqLab, Erlangen), 16,5 µl A. bidest., 1 µl Primer F (10 mM), 1 µl Primer R (10 mM), 2,5 µl dNTP-Mix (jeweils 2 mM dATP, dCTP, dGTP, dTTP), 1 µl Template und 0,5 µl Optimase-Polymerase (2,5 U) eingesetzt. Die Amplifikationen erfolgten im Gene Amp® PCR System 9700 Thermocycler (Applied Biosystems) mit dem folgendem Programm:

- Anfangs Denaturierung: 96 °C, 5min
  - Denaturierung: 96 °C, 30 s
  - Annealing: 55 °C, 45 s
  - Elongation: 72 °C, 60 s
  - Abschluss Elongation: 72 °C, 5 min
  - Kühlung: 8 °C
- } 30 Zyklen jeweils

## 3.4 DNA Aufreinigung der PCR Produkte

Um die Amplifikate von Resten der DNA-Polymerase, überschüssigen dNTPs, Salzen und Puffern zu trennen, erfolgte die Aufreinigung der PCR-Produkte mit dem QIAquick PCR Purifikation Kit von Qiagen. Das Prinzip beruht auf den spezifischen Bindungseigenschaften von Nukleinsäuren (100 bp bis 10 kb) an Silikatsäulen unter Hochsalzbedingungen und deren Elution unter niederosmolaren Bedingungen. Die gereinigten PCR-Fragmente wurden anschließend in 100 µl A. bidest. eluiert und ebenfalls im 1,5 % Agarosegel mit 1 x TBE als Elektrophoresepuffer aufgetrennt und überprüft.

### 3.5 Sequenzierungen

Die klassische Sanger-Sequenzierung wurde im Rahmen dieser Arbeit verwendet zur:

- Identifikation von SNPs in 115 verschiedenen Gen-Fragmenten (Loci)
- SNP-Analyse in den Genen *lukS/F-PV*, zur Bestimmung der von O'Hara *et al.* (2009) beschriebenen PVL-Varianten
- *spa*-Typisierung
- Multilokus Sequenz Typisierung (MLST)

Des Weiteren wurde die *Next Generation Sequencing*-(NGS)-Technologie der 454-Pyrosequenzierung benutzt, um für zehn der 154 untersuchten Isolate eine vollständige Genomsequenzierung durchzuführen.

#### Klassische Sanger-Sequenzierung

Die Identifikation der anhand der dHPLC detektierten SNPs erfolgte durch die Sequenzierung der spezifischen Loci nach dem Prinzip des Didesoxynukleotid-(ddNTPs)-Kettenabbruchverfahrens von Sanger [136]. Die DNA-Sequenzierungen wurden mit dem BigDye® Terminator v1.1 Cycle Sequencing Kit (GE Healthcare) nach Angaben des Herstellers durchgeführt. Für den Sequenzierungsansatz wurde das jeweilige PCR-Produkt mit dem QIAquick PCR Purification Kit aufgereinigt und anschließend für die Sequenzierung eingesetzt. Die Sequenzierungsansätze setzen sich wie folgt zusammen:

- Template: 0,25 µl
- A.bidest.: 8,25 µl
- BigDye 3.1.: 1 µl
- Primer: 0,5 µl

Die Reaktion erfolgte im Thermocycler DNA Engine Dyad™ Peltier Thermal Cycler (Biozym) mit folgendem Programm:

- |  |   |                   |
|--|---|-------------------|
| <ul style="list-style-type: none"> <li>• Denaturierung: 96 °C, 10 s</li> <li>• Primer-Annealing: 50 °C, 5 s</li> <li>• Elongation: 60 °C, 4 min</li> </ul> | } | 25 Zyklen jeweils |
| <ul style="list-style-type: none"> <li>• Kühlung: 4 °C</li> </ul>  |   |                   |

Die Auftrennung der Reaktionsprodukte mittels Kapillargelelektrophorese nach der Sequenzlänge und die Messung der vier verschiedenen Fluoreszenzfarben des mitgelieferten Big Dye's, wurde durch den automatisierten ABI Prism 377 DNA Sequencer im Robert Koch-Institut (Standort: Berlin) durchgeführt. Die Auswertung

der Sequenzierdaten erfolgte primär mit Hilfe der Software LaserGene (Version 7.2.1.(1),410) und für die weitere Datenauswertung mit BioNumerics (Version 5.10).

### Identifikation von Pantone-Valentine-Leukozidin-Genotypen

Damit die von O'Hara *et al.* 2008 und Brown *et al.* 2012 vorgeschlagene Einteilung in verschiedene PVL-Genotypen erfolgen konnte, wurden alle PVL-positiv getesteten CC121-Isolate auf vorhandene Sequenzpolymorphismen in dem 1726 bp großen PVL-Fragment untersucht [113, 13]. O'Hara *et al.* unterschieden Isolate mit PVL aufgrund einer nicht-synonymen Mutation an Position 527, bei der es zum Aminosäureaustausch (Arginin(R) → Histidin(H)) kommt, in die zwei PVL-Genotypen R und H [113]. Diese können aufgrund von weiteren Sequenzpolymorphismen in die PVL-Genotypen R1, R2, H1, H2 und H3 subklassifiziert werden [113, 13]. Zu Beginn der Arbeit waren 13 verschiedene Sequenzpolymorphismen in den *lukS/F-PV* Genen bekannt [113, 13]. Diese wurden den unterschiedlichen PVL-Genotypen zugeordnet [113, 13]. Die für die Sequenzierung verwendeten Primer entsprechen den von O'Hara *et al.* (2008) publizierten und sind in Tabelle A.4 aufgeführt [113].

### *S. aureus*-Protein-A-Typisierung (*spa*-Typisierung)

Eine molekulare Typisierungsmethode von *S. aureus* ist die Analyse von Polymorphismen in der variablen Region X des Protein A Gens *spa*. Protein A repräsentiert einen der vielen Pathogenitätsfaktoren in *S. aureus*, der vor der Opsonierung durch Antikörper schützt, indem es an deren Fc-Fragment bindet. Innerhalb des Gens *spa* befindet sich eine variable Region, die sogenannte X-Region. Diese X-Region ist aus kurzen DNA-*Repeats* aufgebaut und durch Duplikationen, Deletionen oder Punktmutationen innerhalb einer Population sehr variabel. Die Differenzierung dieser *Repeats* definiert den jeweiligen *spa*-Typ eines Isolates.

### Multilokus-Sequenz-Typisierung (MLST)

Die MLST bietet ein exzellentes Werkzeug zur Erforschung der klonalen Populationsstruktur von *S. aureus* und ermöglicht die Ableitung phylogenetischer Beziehungen. Bei der MLST werden sieben Haushaltsgene des Primärstoffwechsels sequenziert und auf Einzelnukleotidpolymorphismen untersucht. Bei den PCR-Fragmenten handelt es sich um die Haushaltsgene der Carboxyltransferase *arcC*, der Shikimat-Dehydrogenase *aroE*, der Glycerolkinase *glpF*, Guanylatkinase *gmk*, der Phosphat-acetyltransferase *pta*, der Triosephosphatisomerase *tpi* sowie um die Acetylcoenzym A Acetyltransferase *yqiL* [33]. Dadurch wird es möglich, die untersuchten Isolate spezifischen Allelprofilen bzw. einem charakteristischen Sequenztyp (ST) zuzuordnen. Die Zuordnung resultiert aus den unterschiedlichen Sequenzen in den jeweiligen Haushaltsgenen der einzelnen Isolate. Die *spa*-Typisierung und die MLST-Methode

wurde hauptsächlich vom technischen Personal des Robert Koch-Institutes übernommen und im Zuge der routinemäßigen Aufgaben des NRZs für Staphylokokken durchgeführt.

### **Genomsequenzierung mit der 454-Pyrosequenzierung**

Das Verfahren der 454-Pyrosequenzierung wurde von der Firma 454 Life Sciences, heute Roche Diagnostics die durch Jonathan Marc Rothberg gegründet wurde, kommerzialisiert. Die Methode basiert auf der bereits Anfang der 1990er Jahre von Mostafa Ronaghi und Pal Nyren erdachten Pyrosequenzierung und wurde durch die von Rothberg gegründete Firma miniaturisiert [130, 92]. Im ersten Schritt der Methode wird die zu sequenzierende DNA physikalisch zerlegt, die dabei entstehenden doppelsträngigen DNA-Bruchstücke (300 - 800 bp) der Größe nach sortiert und mit zwei verschiedenen Oligonukleotid-Adaptoren ligiert. Diese Adapter dienen den Primern als Andockstelle für die Strangverlängerung. Einer der Adapter ist biotinyliert, damit die Aussortierung einzelsträngiger Fragmente für die weitere Sequenzierung ermöglicht wird. Anschließend werden einzelne Mikroreaktoren gebildet, indem die Matrizen-DNA in einzelne Fragmente aufgeteilt und an ca. 28 µm großen sog. *DNA-Capture-Beads* in einer Emulsion immobilisiert werden. In diesen Mikroreaktoren finden die Amplifikationen der Einzelstränge statt. An die *DNA-Capture-Beads* kovalent gebunden sind komplementäre Primer, welche die PCR-Produkte auf der Oberfläche der Beads immobilisieren. Nach Abschluss der Emulsions-PCR ist die Oberfläche der Kugeln vollständig mit den gleichen Sequenzabschnitten überzogen. Nach der Emulsions-PCR werden die Mikroreaktoren aufgebrochen und jedes mit DNA beladene *DNA-Capture-Bead* wird anschließend in eine einzelne Vertiefung auf einer PicoTiter-Platte gegeben, wo die eigentliche Pyrosequenzierung stattfindet. Hierbei wird über eine Enzymkaskade Licht aus dem abgespaltenen anorganischen Pyrophosphat (PPi) erzeugt. Dies geschieht, wenn jeweils eines der vier Nukleotide über die Vertiefungen gespült wird und der Einbau eines passenden Nukleotids durch die DNA-Polymerase gelingt. Das abgespaltene anorganische PPi wird von der ATP-Sulfurylase in Gegenwart von Adenosin-5'-phosphosulfat (APS) zur Herstellung von Adenosintriphosphat (ATP) benötigt, das wiederum Substrat der Luciferase bei der Umwandlung von Luciferin zu Oxiluciferin ist. Bei dieser Umwandlung wird Energie in Form eines Lichtblitzes abgegeben. Die Intensität der erzeugten Lichtsignale ist proportional zu der Anzahl der eingebauten Nukleotide.

### 3.6 SNP-Analyse mittels dHPLC

Die Detektion von Einzelnukleotid-Polymorphismen (engl. *Single Nucleotide Polymorphism*, SNP) erfolgte nicht primär durch DNA-Sequenzierungen, sondern, wie bereits in einigen anderen Studien gezeigt, durch die Heteroduplex-Analyse mittels eines spezialisierten Hochdruckchromatographie-Verfahrens (denaturierende Hochdruckchromatographie, dHPLC) [111, 131]. Mit dieser Methode konnten über einhundert DNA-Fragmente aus allen untersuchten *S. aureus*-Isolaten miteinander verglichen werden. Dabei wird das zu untersuchende DNA-Fragment eines spezifischen Genabschnittes (Lokus) durch PCR amplifiziert (siehe Abschnitt 3.3 auf Seite 25), anschließend mit dem PCR-Produkt einer Referenz-DNA gemischt und mittels dHPLC chromatographisch auf Heteroduplexe zum Referenzstamm untersucht. Heteroduplexe entsprechen doppelsträngigen Nukleinsäuremolekülen, die aus Einzelsträngen unterschiedlicher Herkunft zusammengesetzt sind. Da aufgrund des zunehmenden Anteils ungepaarter Regionen ein Heteroduplex instabiler wird, kann die Passgenauigkeit der Basenpaarung durch die Analyse von Schmelzkurven bestimmt werden. In einem mittels PCR amplifizierten Genabschnitt kann so überprüft werden, ob im Vergleich zu einer Referenz-DNA Sequenzunterschiede vorliegen, die dann gezielt durch Sequenzierung evaluiert werden. Es wurden ausschließlich Haushaltsgene verwendet, da diese für phylogenetische Analysen besonders gut geeignet sind. Bevor die Proben in der dHPLC untersucht werden konnten, erfolgte ein Denaturierungsschritt bei 96 °C für 5 min. Anschließend wurden die Proben in 12 min langsam auf 12 °C abgekühlt, um wieder eine Renaturierung der vorher entstandenen Einzelstränge zu Doppelsträngen zu ermöglichen. Dabei konnten die Einzelstränge einzelner Isolate mit den Einzelsträngen der Referenz-DNA kombinieren. Die entstandenen Hetero- und Homoduplices wurden mittels eines Autosamplers auf die DNASep HT Säule (Transgenomic, Inc., Omaha, USA) injiziert und an diese gebunden. Die Trennung erfolgte bei einer spezifischen Temperatur und Flussrate, die im Vorfeld für jedes Primerpaar als Methode definiert wurde (siehe Tabelle A.3). Der erforderliche Temperatur- und Flussratengradient wurde mit den zugehörigen WAVE-Puffern: WAVE-Puffer A (0,1 M Triethylammoniumacetat-(TEAA) pH 7,0) und WAVE-Puffer B (0,1 M TEAA, 25 % Acetonitril) aufgebaut. Die Detektion der getrennten Genfragmente erfolgt anschließend photometrisch mit einem UV-Detektor. Die spezielle Auswertung und Darstellung der resultierenden Chromatogramme wurde mit Hilfe der gerätespezifischen Navigator Software durchgeführt. Für die vorliegende Arbeit wurde das Isolat 98-01618 als Referenzisolat ausgewählt. Um das Kosten- und Zeitmanagement möglichst effektiv zu gestalten, wurden die Isolate zunächst in Multiplexansätzen untersucht.

**Multiplexansätze.** Bei den Multiplexansätzen handelte es sich um DNA-Gemische, bei denen jeweils die DNA von neun bis zehn Isolaten mit der DNA des Referenzstammes gemischt wurden. Anschließend wurden die einzelnen Multiplexan-

sätze, die sich aus dem DNA-Gemisch verschiedener Isolate zusammensetzen, für die einzelnen Loci (Tab.A.3) amplifiziert und mit Hilfe der dHPLC auf Heteroduplexe zum Referenzstamm untersucht. Kam es bei der Auswertung des spezifischen Chromatogramms für einen untersuchten Locus zu einer Detektion von Heteroduplexen, dann wurden Simplex-Analysen mit den entsprechenden Isolaten der Multiplexansätze und der Referenz wiederholt.

**Simplexansätze.** Bei den Simplexansätzen wurde im Gegensatz zu den Multiplexansätzen nicht die vollständige DNA mehrerer Isolate mit der Referenz DNA gemischt und untersucht, sondern nur das PCR-Produkt eines Isolates mit dem entsprechenden Produkt des Referenzstammes für einen spezifischen Locus. Kam es bei den Multiplexansätzen aufgrund von Sequenzvariabilitäten neben der Reassoziationsbildung identischer Einzelstränge zur Ausbildung von Heteroduplexen von Strängen unterschiedlicher Allele, dann wurden die Isolate dieses Multiplexansatzes erneut einzeln analysiert. Die entsprechenden PCR-Amplifikate für den jeweiligen Locus wurden anschließend mit den PCR-Amplifikaten der Referenz gemischt und analysiert. Detektierte Sequenzunterschiede in den Isolaten wurden anschließend durch gezielte DNA Sequenzierungen validiert.

# 4. Bioinformatische Auswertung von Sequenzdaten

## 4.1 SNP-Analyse

Die Auswertung der Sequenzdaten, die mit Hilfe der klassischen Sanger-Methode generiert wurden und zur Detektion der SNPs dienten, erfolgte in erster Linie mit der Software DNASTAR Lasergene 8. Die Sequenzen der untersuchten Loci wurden als Elektropherogramme übereinandergelegt und auf Sequenzunterschiede zum Referenzstamm 98-01618 untersucht. Anschließend wurden die Sequenzen der Loci mit den vermeintlichen SNPs mit bereits publizierten *S. aureus*-Genomen in der NCBI-Gendatenbank verglichen. Mit Hilfe des Suchalgorithmus BLAST (*basic local alignment search tool*) und dem resultierenden Vergleich bereits publizierter *S. aureus*-Genome konnte ermittelt werden, ob die identifizierten Sequenzunterschiede als ursprünglich (engl. *ancestral*) oder als erworben (engl. *derived*) zu bewerten waren. Diese Unterscheidung ist notwendig, um die Phylogenie der untersuchten CC121 Population zu wurzeln. Mit Hilfe der Software BioNumerics v6.5 war es möglich, alle identifizierten SNPs einzeln zu katalogisieren. Dafür wurde eine Datenbank erstellt, in der die Informationen über die jeweiligen Loci und dessen SNPs zusammen mit der zugehörigen Sequenzinformation und der Position der SNPs im *S. aureus*-Genom N315 (Akzessionsnummer:BA000018) gespeichert wurden. Die Qualität der erhaltenen Elektropherogramme und der daraus resultierenden DNA-Sequenzen wurde sowohl von der Software als auch visuell kontrolliert. Bei Bedarf wurden mangelhafte DNA-Sequenzierungen wiederholt. Nach abgeschlossener Qualitätsprüfung wurden die Sequenzen der einzelnen Loci mit Hilfe der Software Artemis ins gleiche Leseraster gebracht. Dies ist für nachfolgende Fragestellung, wie für die Ermittlung von synonymen und nicht-synonymen Basenaustauschen in den betreffenden Loci notwendig. Die untersuchten Sequenzen des Referenzstammes dienten als Matrix für alle anderen zu untersuchten Isolate. Die im Anschluss an die SNP-Analyse nur noch um die Isolat-spezifischen Mutationen in den entsprechenden Isolaten modifiziert werden mussten. Das daraus resultierende Sequenzalignment aller untersuchten Isolate stellt die Grundlage für anschließende phylogenetische Untersuchungen dar. *Minimum-Spanning-Trees* (MSTs) wurden mit der Software BioNumerics v6.5

erstellt. Dafür wurde separat eine weitere Datenbank in BioNumerics kreiert. In dieser Datenbank wurden jedem Isolat seine Mutationen mit Hilfe eines Binärcodes (1 entspricht erworbenes Merkmal; 0 ursprüngliches Merkmal) zugewiesen und anschließend wurde mit Hilfe des implementierten Algorithmus der MST generiert.

## 4.2 Umgang und Bearbeitung von Genomsequenzen

Wie bereits im Abschnitt 3.5 auf Seite 28 beschrieben wurde, beruhen die Rohdaten der 454-Pyrosequenzierung auf der Detektion von Lichtblitzen mittels einer Kamera. Daraus resultieren Bilddateien mit Größendimensionen von mehreren Gigabyte pro Genom. Daher wurden die Bilddaten in ein benutzerfreundliches Format (sogenannte *Reads* (Abfolge von Nukleotiden)) umgewandelt. Bei der Verarbeitung der Sequenzdaten kann in *de novo* Assemblierung und das Mapping gegen eine Referenzsequenz unterschieden werden. Bei der *de novo* Assemblierung kommt es zur Assemblierung der *Reads* (in den überlappenden Bereichen der einzelnen Reads) unabhängig von unterstützenden Informationen eines Referenz-Genoms. Vorteil dieser Methode ist, dass auch unbekannte Sequenzen zusammengesetzt werden können und dass Restrukturierungen, wie Insertionen oder Deletionen, zuverlässig detektiert werden. Beim Mapping der *Reads* gegen eine bekannte Referenzsequenz werden Basenaustausche detektiert, aber nicht-homologe Bereiche bleiben unerkant.

### Verarbeitung der *Reads*

Die Rohdaten der Sequenzierung, bestehen aus einer Fasta-Datei (\*.fasta, \*.fas) in der die reinen Sequenzinformationen eines jeden *Reads* gespeichert sind und aus einer Qual-Datei (\*.qual) in der die zugehörigen Qualitätswerte für jedes einzelne Nukleotid gespeichert sind. Diese sogenannten PHRED-Qualitätswerte wurden schon mit der Sanger-Sequenzierung eingeführt und haben sich anschließend auch für andere Sequenziermethoden zur Berechnung aussagekräftiger Qualitätswerte pro Base durchgesetzt [35, 36]. Der PHRED-Qualitätswert  $Q$  eines Nukleotids ist folgendermaßen definiert:

$$Q_{PHRED} = -10 * \log_{10}(P_e) \quad (4.1)$$

Für die weitere Bearbeitung der Rohdaten, insbesondere für das Mapping, mussten die beiden Dateien in einer einzigen zusammengefasst werden. Für diese kombinierte Datei hat sich das FASTQ-Format als Standard durchgesetzt. Im FASTQ-Format werden die Werte im ASCII-Alphabet kodiert. Pro Sequenz nutzt das Format vier Zeilen. Die erste beginnt mit einem „@“ und wird meist durch einen spezifisch für die Sequenz vorgesehenen Code gefolgt, ähnlich der Beschreibung für das Fasta-



Format. Die zweite Zeile enthält die kodierte Sequenzinformation, die dritte beginnt optional mit einem „+“ und die letzte Zeile kodiert die jeweiligen Qualitätswerte für alle enthaltenen Nukleotide aus Zeile 2. Mittlerweile gibt es mehrere FASTQ-Varianten, je nachdem welche Sequenzierungstechnologie bevorzugt wird. Dies ist ein Beispiel für die Standardisierungsschwierigkeiten bei der Hochdurchsatzsequenzierung und die daraus resultierenden Schwierigkeiten Qualitätswerte unterschiedlicher Technologien miteinander vergleichen zu können. Die Konvertierung der Fasta und Qual-Dateien in das FASTQ-Format konnte mit einem frei verfügbaren Perl-Skript „*FastaTOFastq.pl*“ realisiert werden.

## Mapping

Für das Mapping der Sequenz-*Reads* aus der 454-Pyrosequenzierung wurde die Linux Software SSAHA2 verwendet [104]. Als Referenzsequenz wurde das Genom des Isolates N315 ausgewählt. Beim Mapping wurde SSAHA2 mit den *default*-Einstellungen ausgeführt. SSAHA2 generiert als Outputformat eine SAM-Datei \*.sam, die mit Hilfe einer weiteren Software Plattform SAMtools v0.1.12a zum textbasierten Pileup-Format \*.pileup generiert wurde [79]. Dies ist wichtig, da sich die SAM-Datei nicht zur leichten Weiterverarbeitung eignet. Anschließend diente ein Filter dazu, Sequenzierungsfehler zu entfernen und eine Qualitätsgeprüfte Sequenz zu erhalten. Dies wurde mit einem Python-Skript „*pileupFixSSAHA2.py*“ realisiert, das Frau Dr. Janina Dordel (ehemalige Doktorandin im Robert Koch-Institut) im Rahmen ihrer Dissertation entwickelt hat [31]. Dieses Skript filtert die Basen mit einem definierten Qualitätsgrenzwert heraus und ersetzt diese durch "N"(unspezifische Nukleobase). Im Rahmen dieser Arbeit wurden für die Einstellung folgende Werte angenommen: 30 für die Konsensus-Qualität, 5 für die *Coverage* und 30 für die Mapping Qualität. Nach erfolgreicher Filterung wurden zwei Dateien generiert: eine \*.fix Datei, die dem Pendant der \*.pileup Datei entspricht, nur dass sie die veränderten Basen enthält und eine \*.fas Datei, die die entsprechende Sequenz im Fasta-Format beinhaltet. Für die Klärung der Phylogenie sollte nur das Kerngenom der sequenzierten Isolate einbezogen werden. Daher wurden alle mobilen genetischen Elemente und Wiederholungssequenzen aus der zuvor qualitätsgefilterten Sequenz entfernt. Diese Aufgabe wurde mit einem weiteren Perl-Skript „*RegionRemover.pl*“ durchgeführt, das ebenfalls von Frau Dr. Janina Dordel stammt [31]. Als Input-Dateien dienten dabei zum einen die vom „*pileupFixSSAHA2.py*“ erstellte Fasta-Datei und eine vorher erstellte Textdatei positionlist.txt, in der sämtliche Start- und Endpositionen der mobilen genetischen Elemente im Referenzgenom gespeichert sind. Das Skript entfernte alle in der Textdatei positionlist.txt gespeicherten Regionen aus der gemappten Sequenz und erstellte eine neue Fasta-Datei \*.cleaned-sequence.fasta, die die entsprechenden Konsensus-Sequenzen enthielt.

## 4.3 Phylogenetik von CC121

Nach dem Mapping, der Qualitätsfilterung und der Entfernung der mobilen genetischen Elemente sowie der Repeatregionen, bietet die Software MEGA5 die Möglichkeit die vorliegenden qualitätsgeprüften Sequenzen im FASTA-Format (\*.cleaned-sequence.fasta) schnell und einfach in andere Formate zu konvertieren oder weiter zu verarbeiten. Mit dem implementierten Programm ClustalW konnte ein multiples Sequenzalignment erstellt werden. Anschließend wurde ein SNP-Alignment extrahiert. Dieses ist notwendig, um den Rechenaufwand bei den anschließenden phylogenetischen Analysen zu minimieren. Vor der Erstellung des SNP-Alignments erfolgte eine zusätzliche manuelle Qualitätskontrolle aller vorliegenden SNPs im Alignment. Dabei wurde überprüft, ob sich SNPs in homopolymeren Sequenzabschnitten befanden, in unmittelbarer Umgebung „ $< 50bp$ “ von Indels auftauchten oder ungewöhnliche SNP-Ansammlungen vorlagen, die auf mögliche Sequenzierfehler hinweisen. Diese Sequenzbereiche wurden entsprechend ausgeschlossen. Erst nach dieser zusätzlichen Qualitätsprüfung wurde ein SNP-Alignment erstellt, das für die nachfolgenden phylogenetischen Analysen verwendet werden konnte. Eine weitere wichtige Funktion von MEGA5 war es, die Daten in das NEXUS-Format (\*.nxs; \*.nex) umzuwandeln, das wiederum für andere Programme benötigt wurde.

### 4.3.1 Phylogenie mit der *Maximum-Likelihood*-Methode

Die *Maximum-Likelihood*-(ML)-Methode ist eine phylogenetische Methode, bei der ein Stammbaum ermittelt wird, der unter einem gegebenen Modell der Sequenz evolution die Wahrscheinlichkeit (*likelihood*) der gegebenen Daten maximiert. Es gehört zu den Zwei-Schritt-Verfahren, da während der Suche zuerst viele Stammbäume generiert werden, die dann wiederum miteinander verglichen werden, bis ein Baum die vorgegebenen Optimalitätskriterien am besten erfüllt. Die ML-Methode wurde mit Hilfe der Software PhyML durchgeführt, die in Seaview implementiert war.

Durch das Einführen einer Außengruppe (engl. *outgroup*) konnte die Wurzel „*Root*“ des ML-Baumes bestimmt werden. Als Außengruppe wurde das *S. aureus*-Genom N315 verwendet. N315 gehört zum Klonalen Komplex 5 und ist somit nicht unmittelbar verwandt mit den untersuchten Isolatens des Klonalen Komplexes 121. Für die Rekonstruktion der Außengruppe wurde das zur Verfügung stehende Genom von N315 ebenfalls in eine konkatenierte Sequenz mit einer Gesamtlänge von 46,811 bp transformiert und mit den 154 zu untersuchenden *S. aureus*-Isolaten aligniert. Dies konnte durch die entsprechenden Informationen über die zu untersuchenden Loci und mit den jeweiligen Primerpositionen im Genom realisiert werden. Zusätzlich wurden noch *Bootstrap*-Werte berechnet. *Bootstrap*-Werte geben eine Schätzung über die Verlässlichkeit eines Knotens in einem Stammbaum wieder [74]. Für die folgenden

Berechnungen wurden 1000 Wiederholungen für die *Bootstrap*-Analyse eingestellt. Die ML-Methode wurde neben den Berechnungen der vollständigen Genomsequenzen auch für das durch die Software BioNumerics erstellte Sequenzalignment von 46.811 bp verwendet (siehe Abschnitt 4.1 auf Seite 31).

### 4.3.2 Phylogenie mittels Bayes'sche Analysen

Die Bayesianische Statistik sucht die phylogenetischen Bäume mit der höchsten *posterior probability*. Der Unterschied zum ML-Verfahren ist, dass die Wahrscheinlichkeit der Hypothese anhand der Daten berechnet wird und nicht die Wahrscheinlichkeit der Daten anhand der Hypothese [74]. Ziel dieser Analysen sind Stammbäume mit den höchsten *posterior probability*-Werten. Die Bayesianische Analyse zur Ermittlung der Phylogenie von CC121 wurde mit Hilfe der Software MrBayes realisiert. Folgende Eingaben wurden verwendet um die Bayes'sche Analyse in MrBayes zu starten.

Mit dem Befehl **lset nst= 6 rates= invgamma** wird das Substitutionsmodell definiert. Dieses Substitutionsmodell entspricht einem *General Time Reversible Model* (GTR) mit einem Anteil invarianter Positionen und einer Gamma-Verteilung (GTR-G+I). Der zweite Befehl legt die verschiedenen *Priors* fest:

```
prset revmatpr= dirichlet(1,1,1,1,1,1) statefreqpr= dirichlet(1,1,1,1,1,1)
shapepr=uniform(0.1,50) pinvarpr=uniform(0,1).
```

Hierbei handelt es sich um die *default*-Einstellung von MrBayes. Der letzte Befehl startet die Analyse, bei dem der Metropolis-coupled Markov Chain Monte Carlo (MCMCMC)-Algorithmus angewendet wird. Der Befehl dafür lautet:

```
mcmc ngen= 15000000 nchains= 4 samplefreq= 15000 printfreq= 100000
savebrlens=yes.
```

Bei der Analyse laufen zwei parallele Berechnungen mit je vier Ketten, bei der die Markov-Kette mit 15.000.000 Generationen durchlaufen wird. Mit **samplefreq= 15000** wird erreicht, dass die Bäume jeder 15.000sten Generation gesammelt und mit dem Befehl **savebrlens=yes** samt der dazugehörigen Astlänge in eine Datei gespeichert werden. Im Anschluss an die Analyse, wird ein Konsensusbaum (*Majority Rule*-Stammbaum) aus den 1000 gespeicherten Bäumen erstellt. Der dafür verwendete Befehl ist *sumt*. Als Burn-in wurden dabei die ersten 250 Bäume aus der Analyse verworfen und mit dem Befehl **sumt burnin=250** an die Software MrBayes weitergegeben.

Grund dafür ist, dass die Ketten eine sogenannte Anlaufstrecke benötigen. Bevor sie Bäume mit optimalen *Likelihoods* finden, gibt es bei jeder Analyse eine sogenannte *Burn-in* Phase in der sich Bäume mit mehr oder weniger suboptimalen *Likelihoods* befinden. In der Regel werden daher die ersten 10 - 15 % der gefilterten Bäume verworfen und als Burn-in deklariert [74].

### 4.3.3 Datierung und Substitutionsraten mittels BEAST

BEAST (englisch: *Bayesian Evolutionary Analysis Sampling Trees*) ist eine Software, entwickelt von Alexei Drummond und Andrew Rambaut, die zur Analyse von molekularen Sequenzen mittels Bayesverfahren unter Verwendung von MCMC-Methoden verwendet wird [32]. Für die in der Arbeit untersuchten Fragestellungen wurde BEAST zur Analyse von Substitutionsraten und zur molekularen Datierung der CC121-Phylogenie verwendet. Mit der Software BEAST wird gleich eine ganze Reihe zusätzlicher Softwarepakete bereitgestellt, die zum Teil zur Verarbeitung des *Outputs* sowie zur Generierung des *Inputs* verwendet werden können. Da BEAST auf dem XML-Dateiformat zur Beschreibung der Modelleinstellungen basiert, muss die zuvor mit der Software MEGA5 generierte NEXUS-Datei mit dem Sequenzalignment der 154 untersuchten Isolate durch die begleitende Software namens BEAUTi zuerst generiert werden. Auf der benutzerfreundlichen Oberfläche von BEAUTi müssen die Einstellungen zum Modell der Sequenzevolution, zum Modell der molekularen Uhr und die entsprechenden *Priors* eingestellt werden. Für das Substitutionsmodell wurde das Modell von Hasegawa-Kishino-Yano (HKY-Modell) verwendet, als Modell der molekularen Uhr wurde sowohl das *strict clock model* als auch das *relaxed lognormal* gewählt. Als *Priors* wurden die vordefinierten *default*-Einstellungen belassen. Die entsprechenden Generationen, wie lange die Markov-Kette laufen soll, mussten ebenfalls definiert werden. Da die Analyse mit BEAST sehr rechenintensiv ist und der Zeitaufwand hoch, wurde im ersten Lauf eine geringe Anzahl an Generationen gewählt. Erst im Anschluss wurden die Generationen so approximiert, dass nach der Berechnung des zugrunde liegenden MCMCMC-Algorithmus der Parameter für die *effective sample size* (ESS) den Wert von 200 überstieg. Erst dann konnte sicher davon ausgegangen werden, dass die Kette lange genug gelaufen war und die Ergebnisse zuverlässig erschienen [74]. Der im Rahmen dieser Arbeit untersuchte Datensatz mit dem generierten SNP-Alignment wurde mit  $2 * 10^8$  Generationen durchlaufen und die Bäume jeder  $2 * 10^5$ sten Generation mit Astlängen in einer TEXT-Datei (\*.trees.txt) gespeichert. Auf diese Weise konnten für jede Analyse 1000 Bäume abgespeichert werden. Wie auch schon bei der Bayes-sche Analyse beschrieben, wurde auch hier eine *Burn-in* Phase von 10 % angenommen und die ersten 250 Bäume für die Analyse verworfen. Anschließend wurde mit der Software Tracer v1.5 überprüft, ob sich die Ketten nach der gewählten Anzahl an Generationen im Equilibrium befanden und eine ESS von mindestens 200 erreicht werden konnte. War dies nicht der Fall, wurde die Analyse mit einer gesteigerten Generationszahl wiederholt.

## 4.4 Korrelation zwischen phänotypischen Merkmalen und der Phylogenie

Um eine signifikante Aussage über die Korrelation von phänotypischen Merkmalen und der Phylogenie treffen zu können, wurde die Software BaTS (***B**ayesian **T**ip-association **S**ignificance testing*) verwendet. BaTS wurde von Joe Parker entwickelt und für Java kompiliert [121]. BaTS beruht auf der Berechnung von drei Statistiken, die die Stärke einer Korrelation von einem bestimmten Merkmal und der zugehörigen Phylogenie beschreibt. Zum einen wird der *Association Index* (AI) berechnet, zum anderen der *parsimony score*-(PS)-Wert und der neu mit der Software eingeführte *maximum exclusive single-state clade size*-(MC)-Wert [135]. Der AI ist definiert durch die Summe aller internen Knoten in der bestehenden Phylogenie und wird durch folgende Formel beschrieben:

$$AI = \sum_{i=1}^k \frac{1 - f_i}{2^{m_i - 1}} \quad (4.2)$$

Dabei steht  $k$  für die Anzahl der internen Knoten  $i$ ,  $f_i$  steht für die Frequenz der am häufigsten vorkommenden Merkmalszustände innerhalb der Äste bezogen auf die gegenüberliegenden Knoten,  $m_i$  ist die Anzahl der Äste entgegengesetzt zu den Knoten  $i$ . Kleine AI-Werte repräsentieren eine starke Assoziation von Phylogenie und betreffendem Merkmal [153].

Die PS-Statistik wird nach dem *Parsimony* Algorithmus von Fitch berechnet [40]. Unter der Annahme, dass der Gewinn oder der Verlust eines Merkmals im allgemeinen nur sehr selten stattfindet, sollte der beobachtete PS-Wert sich umgekehrt proportional zu der Stärke der Korrelation von Phylogenie und betreffendem Merkmal verhalten. Die Werte der PS-Statistik liegen in dem Bereich von  $1 \leq PS \leq n$ , wobei  $n$  die Anzahl der gegebenen Äste ist. Auch bei dieser Statistik repräsentieren kleine PS-Werte eine starke Assoziation von Phylogenie und betreffendem Merkmal [40]. Bei der MC-Statistik geht man von der Hypothese aus, dass eine starke Assoziation von Phylogenie und betreffenden Merkmalszuständen durch große monophyletische Gruppen repräsentiert sein sollten, in denen alle terminalen Äste gleiche Merkmale teilen. Diese Eigenschaft wird durch die sogenannte *monophyletic clade*-(MC)-Größenstatistik für ein bestimmtes Merkmal  $x$  quantifiziert und ist definiert durch:

$$MC(x) = \max_{i=1}^k (m_i I_i) \quad (4.3)$$

Dabei ist  $m_i$  durch die Anzahl der terminalen Äste gegenüber der Knoten  $i$  und  $k$  durch die Anzahl der internen Knoten einschließlich der Wurzel definiert.  $I_i$  hat die Funktion eines Indikators und bekommt den Wert 1, wenn alle Äste abzweigend von den Knoten  $i$  den Merkmalszustand  $x$  besitzen und anderenfalls den Wert 0. Das Ergebnis ist ganzzahlig und bewegt sich in einem Wertebereich von  $1 \leq MC \leq n_x$ ,

$n_x$  wird durch die Anzahl der Äste mit dem Merkmalszustand  $x$  definiert. MC reflektiert die Stärke der Korrelation von Phylogenie und betreffendem Merkmal. Um die statistische Überprüfung der Korrelationen zwischen Phylogenie und betreffenden Merkmalszuständen nicht durch redundante Haplotypen zu verzerren, wurde für die Berechnungen nur jeweils ein Haplotyp berücksichtigt.

## 4.5 Paup\* zur Berechnung des Homoplasie-Indexes (HI)

Bei der Suche nach der besten Phylogenie kann es zu Konflikten kommen. Diese Konflikte entstehen vor allem dann, wenn sich in einem realen Datensatz Merkmale widersprechen. In diesem Fall wird von Homoplasien gesprochen. Das bedeutet, dass identische Merkmalszustände nicht mehr ausschließlich über die Vererbung anhand einer gegebenen Phylogenie beschrieben werden können. Diese Merkmalszustände können durch konvergente Evolution oder durch Rekombination entstehen. Mit Hilfe der Software PAUP\* (englisch: *Phylogenetic Analysis Using Parsimony*), wobei der kleine Stern seit der neuen Version 4 für “*and other methods*,” steht, ist es möglich den Grad an Homoplasien durch den sogenannte HI-Wert zu beschreiben [156]. HI beschreibt das Verhältnis der Transformationsschritte aller Merkmale zu der Anzahl der detektierten homoplastischen SNPs. Bei einem Wert von 1 sind alle Merkmale homoplastisch, wohingegen bei einem Wert von 0 keine Homoplasien vorliegen. Für die Berechnung wurde das erstellte Sequenzalignment aus Abschnitt 4.1 im NEXUS-Format in PAUP\* eingeladen und mit den entsprechenden Eingabe-Kommandos in PAUP\* der HI berechnet. Da der HI keinerlei Informationen über die Position der homoplastischen SNPs im untersuchten Datensatz liefert, mussten diese anschließend durch die Software DnaSP v5.10.01 ermittelt werden [80].

## 4.6 Sequenzevolution $dN/dS$

Im Laufe der Evolution kann es durch auftretende Mutationen zu einer Divergenz zwischen zwei Genen kommen. Da nicht jede dieser Mutationen aufgrund des degenerierten Kodes zu einer Veränderung der kodierten Aminosäure führt, kann zwischen nicht-synonymen und synonymen Substitutionen unterschieden werden. Die Anzahl der nicht-synonymen Substitutionen wird durch  $dN$  und die Anzahl der synonymen Substitutionen durch  $dS$  wiedergegeben. Bei einer synonymen Substitution bleibt die Aminosäure erhalten, während eine nicht-synonymen Substitution eine Veränderung des Aminosäurerests oder gar ein Stopcodon zur Folge hat. Mit der Software DnaSP v5.10.01 war es möglich durch den paarweisen Vergleich zweier Sequenzen die  $dS$ - und  $dN$ -Werte zu ermitteln und durch die implementierte Formel nach Nei & Gojobori (1986) das normalisierte Verhältnis von  $dN$  zu  $dS$  zu berechnen

[102, 80].

$$dN/dS = \frac{(n/N)}{(s/S)} \quad (4.4)$$

Dabei steht  $n$  für die Summe der nicht-synonymen Mutationen,  $N$  für die Anzahl der nicht-synonymen Stellen,  $s$  für die Summe der synonymen SNPs und  $S$  für die Anzahl der synonymen Stellen. Ist das Verhältnis von  $dN/dS \ll 1$ , wird von reinigender Selektion (*purifying selection*) gesprochen. Das bedeutet, dass Mutationen ausselektiert werden um das kodierte Protein in seiner ursprünglichen Funktionalität zu erhalten. Bei einem  $dN/dS > 1$  wird von diversifizierender Selektion (*diversifying selection*) gesprochen, die daraufhin deutet, dass Veränderungen im betreffenden Protein positiv selektiert werden.

## 4.7 Genom-Annotation

Für den Genomvergleich wurde die Software Kodon v3.6.1 (Applied Maths) verwendet. Kodon bietet neben dem einfachen Alignment ganzer Genome, auch eine sehr gute graphische Oberfläche zur Darstellung der Genom-basierten Alignments. Für die *in silico*-Analysen der vollständig sequenzierten Genome wurden zuerst die erstellten *Assemblies* für jedes sequenzierte Isolat mit den vorhandenen *Contigs* in Kodon eingeladen und mit dessen Hilfe auf der Grundlage der bereits in GenBank publizierten *S. aureus* Genome annotiert. Für die Annotation wurden ORFs „*open reading frames*“ ab einer Größe von  $\geq 99$  bp angenommen. In den Referenzgenomen nicht vorkommende Gene, wurden mit der BLAST-Methode mit den in der GenBank enthaltenden Genannotationen verglichen und bei einer Homologie von mindestens 50 % annotiert.

# Teil III

## Ergebnisse



## 5. Ergebnisse

### 5.1 *S. aureus*-Isolate

Alle 154 untersuchten *S. aureus* Bakterienstämme konnten über die beiden Typisierungsmethoden MLST und *spa*-Typisierung dem Klonalen Komplex (englisch: *clonal complex*, CC) 121 (CC121) zugeordnet werden. Mit der *spa*-Typisierung konnten 38 verschiedene *spa*-Typen identifiziert werden. Die MLST zeigte, dass der Großteil der Isolate dem Sequenztyp (ST) 121 entspricht, der sich durch folgendes Allelprofil definiert: 3-6-5-6-2-7-14-5. Darüber hinaus konnten 15 Einzellokusvarianten (engl. *single locus variants*, (SLV)) detektiert werden, die ebenfalls dem klonalen Komplex CC121 angehören (siehe Tabelle A.1). Die Isolate 07-00384, 07-02764, 09-00227, 10-01148 und 10-01153 konnten mit ihrem MLST-Allelprofil 3-6-5-6-2-7-17-19 als ST123 definiert werden. ST123 stellt eine Doppelokusvariante zum ST121 dar und gehört ebenfalls zum CC121. Für 35 der 154 Isolate konnten detaillierte epidemiologische Daten erhoben werden, die Rückschlüsse auf klinische Phänotypen zuließen. Alle *S. aureus*-Isolate und die dazugehörigen Daten sind in Tabelle A.1 aufgeführt.

### 5.2 SNP-Analyse mittels dHPLC

Nachfolgend werden die Ergebnisse der SNP-Analyse mittels dHPLC zusammengefasst. Damit die Populationsstruktur der 154 *S. aureus*-Isolate innerhalb des CC121 umfassend analysiert werden konnten, wurden Isolate aus 27 verschiedenen Ländern über fünf Kontinente verteilt in das Projekt aufgenommen und untersucht (Tab.A.1). Zu Beginn wurden 125 Haushaltsgene in allen 154 *S. aureus*-Isolaten auf Einzelnukleotidpolymorphismen analysiert. Haushaltsgene kodieren für Proteine, die essenziell für grundlegende Stoffwechselvorgänge und dadurch unentbehrlich für die Organismen sind. Dies hat zur Folge, dass diese Gene unter keinem Sequenz-verändernden Selektionsdruck stehen und neutral evolvieren. Einzelne Abschnitte aus den 125 Genen wurden durch PCR amplifiziert. Alle dafür verwendeten Primer wurden aus den Projekten zu ST5, ST22 und ST225 bezogen und zunächst auf ihre Funktionalität bezüglich der CC121-Isolate überprüft [110, 111]. Konnten bestimmte Loci nicht amplifiziert werden oder lag eine unzureichende Qualität der Amplifikate vor,

dann wurden diese Primer ausgeschlossen. Nach abgeschlossener Sondierung wurden Primerpaare für 115 Haushaltsgene ausgewählt. Die bei der PCR entstandenen Genfragmente wurden auf ihre Heterozygotität zum Referenzstamm (Isolat, 98-01618) mithilfe der dHPLC untersucht. Alle für die Arbeit verwendeten Primer können der Tabelle A.3 entnommen werden. Insgesamt umfasste das Screening 46,839 bp. Dies entspricht ca. 1,7 % des Gesamtgenoms von *S. aureus*, das eine durchschnittliche Genomgröße von 2,8 Mbp besitzt. Wie bereits in Abschnitt 3.6 auf Seite 29 beschrieben, wurden die zu bearbeitenden *S. aureus*-Isolate zuerst in Multiplexansätzen untersucht. Kam es zur Detektion von Heteroduplexformationen, konnten die jeweiligen SNPs durch anschließende Simplexreaktionen detektiert und mithilfe der Sequenzierung validiert werden. Die PCR-Produkte hatten eine durchschnittliche Länge von 400 bp. Bei der SNP-basierten Analyse der CC121-Population wurden in knapp 2000 Multiplexreaktionen ca. 550 Heteroduplexe detektiert und in jeweils zehn Simplexreaktionen auf entsprechende SNPs untersucht.

### 5.2.1 Sequenzierung der detektierten SNPs

Kam es in einem mit Hilfe der dHPLC untersuchten Loci zur Identifikation von SNPs wurden diese sequenziert, um die SNPs genauer zu charakterisieren. Die Auswertung der Daten erfolgte mit der Software Lasergene 8. Hierzu wurde ein Sequenzalignment der zu überprüfenden Sequenzen mit der Referenzsequenz des jeweiligen Lokus erstellt und analysiert. Primersequenzen am Anfang und am Ende der entsprechenden Sequenz wurden entfernt. Nach der Identifizierung eines SNPs wurde die entsprechende Sequenz gegen die NCBI-Datenbank geblastet und mit bereits publizierten *S. aureus*-Genomen verglichen.

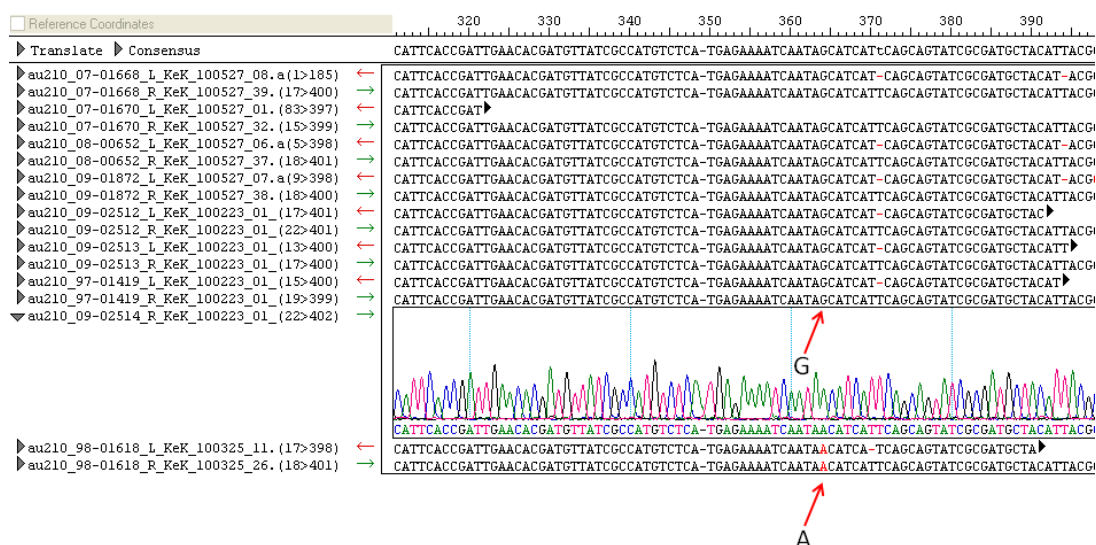


Abbildung 5.1: Alignment von Sequenzdaten verschiedener *S. aureus*-Isolate mit der Referenzsequenz exemplarisch am Locus *au210*

Abbildung 5.1 zeigt exemplarisch das Alignment von acht verschiedenen Isolaten und

dem Referenzisolat für den Locus *au210*. Deutlich im Alignment zu erkennen ist die Transition (Austausch von Purinbasen) an Position 364 bei sieben der ausgewählten Isolate gegenüber dem Referenzstamm (siehe rote Pfeile).

Insgesamt konnten 316 Polymorphismen identifiziert werden, was ungefähr einem SNP pro 0,15 kb entspricht. Von den 316 detektierten Polymorphismen waren 211 Transitionen, 90 Transversionen (Purinbasen gegen Pyrimidinbasen) und 13 Indels. Letztere teilten sich auf in sechs Insertionen und sieben Deletionen in der Größenordnung von 1-12 Nukleotide. Alle identifizierten Polymorphismen und die jeweilige Klassifizierung in Transitionen, Transversionen und Indels können der Tabelle A.2 entnommen werden.

## 5.3 Phylogenetik von CC121

Um die geographische Ausbreitung sowie den Einfluss der genomischen Variabilität auf die Pathogenität zu beleuchten, wurden die Isolate je nach Herkunft, Infektionstyp und Virulenzdeterminanten (ETA, ETB und PVL) mittels eines *Minimum-Spanning-Tree* (MST) und mit dem *Maximum Likelihood*-(ML)-Modell phylogenetisch untersucht.

### 5.3.1 *Minimum-Spanning-Tree*

Mithilfe der Software BioNumerics wurden alle erfassten Sequenzpolymorphismen, die in den 115 Genfragmenten der ausgewählten Haushaltsgene detektiert wurden, zur Erstellung eines MST verwendet. Somit war es möglich, die phylogenetischen Zusammenhänge der untersuchten Isolate darzustellen. Abbildung 5.2 zeigt diesen MST für die untersuchte CC121-Population.

Der dargestellte MST besteht aus 120 verschiedenen Haplotypen (1-119 und R) und gibt die Verwandtschaftsverhältnisse aller Isolate auf Basis der 316 identifizierten Polymorphismen wieder. Die Topologie weist eine sternförmige Struktur auf, bei der alle phylogenetischen Linien aus einem hypothetischen Knoten hervorgehen. Es lassen sich deutlich sechs phylogenetische Sublinien identifizieren. Diese werden jeweils durch mehrere Isolate repräsentiert ( $\geq 10$ ) und sind durch gestrichelte Linien und alphabetisch von A bis F im MST hervorgehoben. In der Tabelle 5.1 sind alle Klade-spezifischen Polymorphismen aufgeführt. Die Position der Wurzel (rot-gestrichelter Kreis) konnte durch den direkten Vergleich aller ermittelten Polymorphismen und den bereits publizierten und vollständig sequenzierten *S. aureus*-Genomen unterschiedlicher klonaler Linien ermittelt werden. Somit konnte die Richtung einer Merkmalsausbildung bestimmt werden, das heißt, es konnte unterschieden werden, ob es sich um ein ursprüngliches oder abgeleitetes Merkmal handelt.

Tabelle 5.1: Klade spezifische SNPs

Klade	SNP-ID	Auswirkung der Mutation	Ursprung	Abgeleitet	Position Genom N315	im N315	ORF in N315
A	au225-1	nicht-synonym	G	T	974479		SA0859
	au234-1	nicht-synonym	C	T	1766929		SA1545
	au346-1	nicht-synonym	G	A	1369964		SA1197
	au319-7	synonym	G	A	402780		SA0344
B	au341-1	nicht-synonym	C	T	1275916		SA1122
	au366-1	synonym	G	A	1701849		SA1493
C	au218-2	nicht-synonym	G	C	592619		SA0507
	au336-3	nicht-synonym	A	T	1042480		SA0917
D	au269-1	nicht-synonym	G	A	2517330		SA2240
	au247-4	synonym	T	C	418717		SA0361
	au392-1	nicht-synonym	T	C	2267501		SA1995
E	au215-2	nicht-synonym	C	T	495953		SA0431
	au218-4	nicht-synonym	A	C	592604		SA0507
	au260-7	nicht-synonym	C	G	1781440		SA1554
	au392-7	nicht-synonym	C	A	2267823		SA1995
F	au210-1	nicht-synonym	C	T	369435		SA0311
	au332-1	synonym	G	A	883327		SA0776
	au355-1	nicht-synonym	C	T	1556430		SA1346

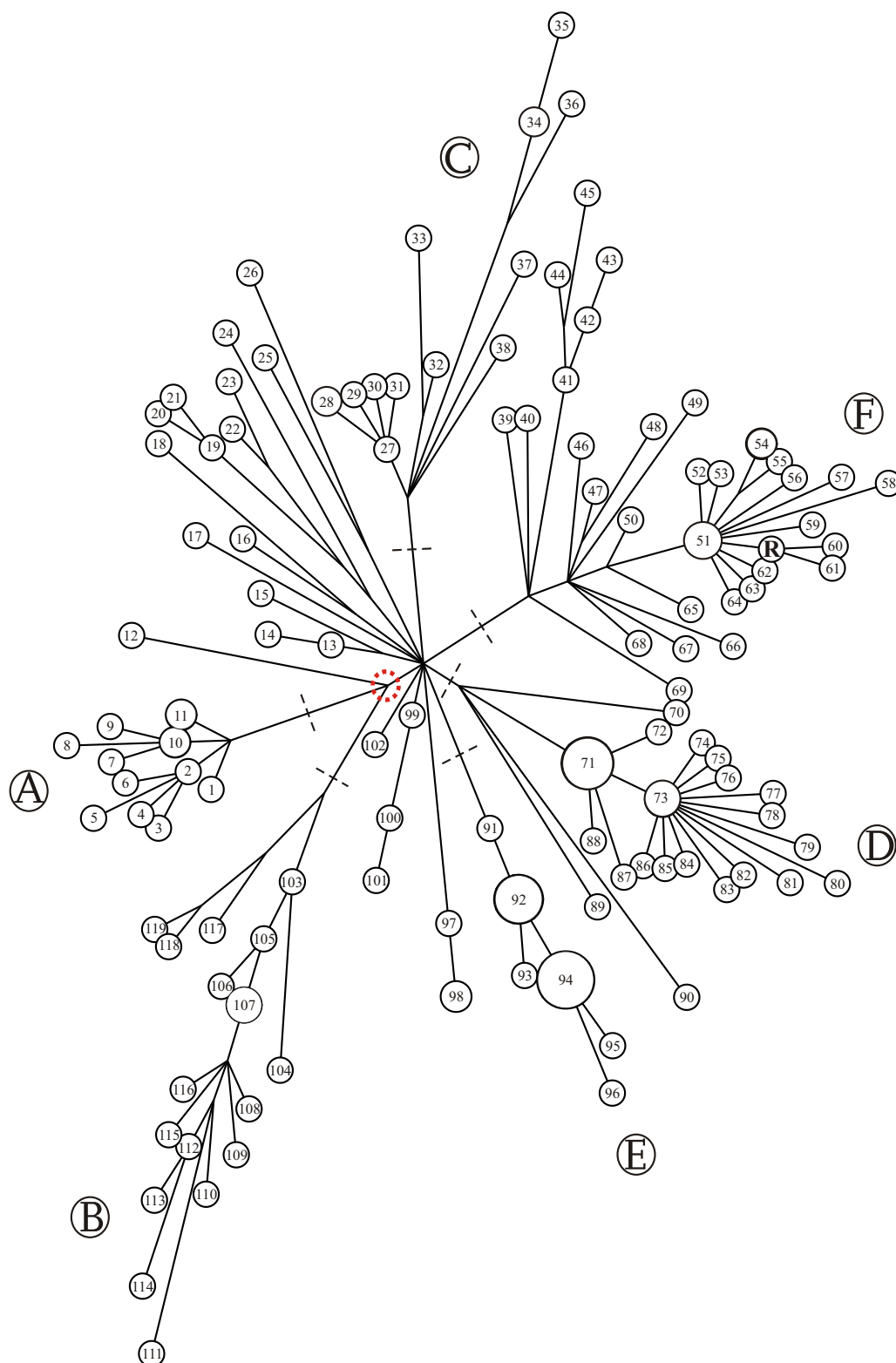


Abbildung 5.2: *Minimum-Spanning Tree* basierend auf SNP-Daten. Haplotypen sind durch Kreise repräsentiert. Die Größe der einzelnen Kreise ist dabei proportional zur Zahl der enthaltenen Isolate. Die Verbindung zweier Haplotypen repräsentiert die Anzahl der jeweiligen Merkmalsaus-tausche in einem Locus, an dem sich die einzelnen Haplotypen und die dazugehörigen Isolate voneinander differenzieren. Der rote gestrichelte Kreis kennzeichnet die Wurzel des Baumes. Die Länge der Verbindungslinien ist dabei proportional zur Anzahl der identifizierten SNPs zwischen zwei Haplotypen. Alle Isolate und die dazugehörigen Haplotypen sind in der Tabelle A.1 vermerkt. Der im MST als “R<sub>1</sub>” gekennzeichnete Haplotyp entspricht dem für die Untersuchungen ausgewählten Referenzstamm, 98-01618.

### 5.3.2 *Maximum-Likelihood-Modell*

Um die Ergebnisse der MST-Analyse vergleichen und die abgeleiteten phylogenetischen Rückschlüsse untermauern zu können, wurde zusätzlich noch die ML-Methode als Zwei-Schritt-Verfahren angewendet. Bei dieser Methode wird die Baumtopologie ausgewählt die als die beste gilt und die Sequenzvariabilität des Gesamtdatensatzes mit der höchsten Wahrscheinlichkeit erklärt. Abbildung 5.3 zeigt das Ergebnis dieser Methode in einem ML-Baum, basierend auf dem multiplen Sequenzalignment der 154 untersuchten *S. aureus*-Isolate und einer Gesamtlänge von 46,811 bp. Wie bereits beim MST in Abbildung 5.4 gezeigt, sind auch im ML-Baum (Abb.5.3) sechs Kladen (A-F) unterscheidbar. Die Position der berechneten Wurzeln im MST in Abb.5.2 und im ML-Baum in Abb.5.3 ist identisch.

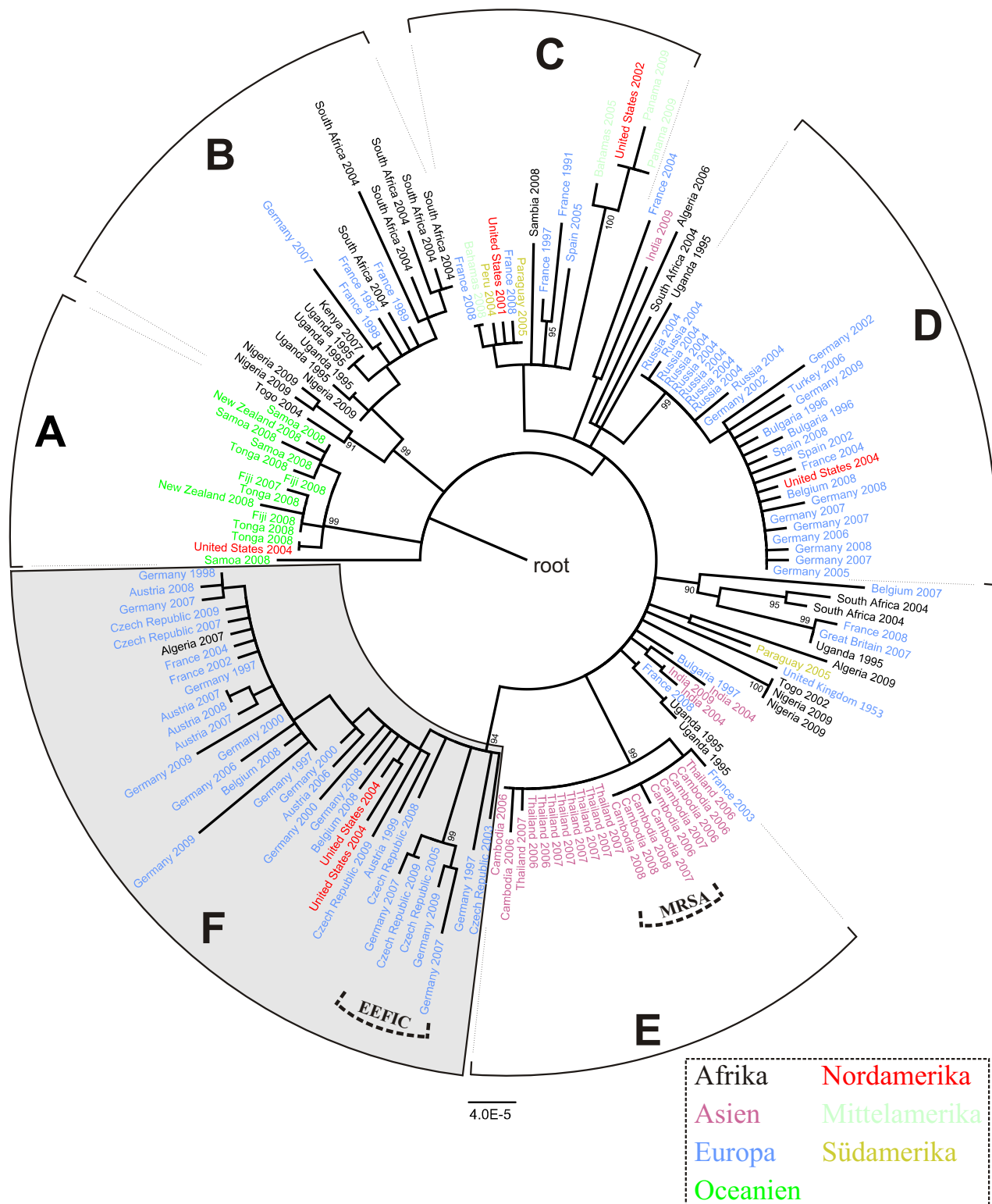


Abbildung 5.3: *Maximum-Likelihood*-Baum beschreibt die Phylogeographie von CC121; berechnet aus konkate-niertem Sequenzalignment von 154 *S. aureus*-Isolaten (46,811 bp) mittels Seaview; Kontinente wurden farbig un-terlegt und nur Bootstrap-Werte über 90 % an den entsprechenden Kanten im Baum eingefügt.

### 5.3.3 Phylogeographie

Die Abbildungen 5.3 und 5.4 geben die Phylogeographie der untersuchten CC121-Population wieder. Im MST sind die verschiedenen Herkunftsländer der untersuchten

*S. aureus*-Isolate mit unterschiedlichen Farben unterlegt. Im ML-Baum repräsentieren die Taxa-Bezeichnungen der terminalen Äste die unterschiedlichen Herkunftsländer der einzelnen Isolate und wurden je nach Kontinent farbig markiert.

Sowohl im MST als auch im ML-Baum ist ein Zusammenhang zwischen Geographie und ermittelter Phylogenie für die untersuchte CC121-Population in den einzelnen Kladen erkennbar. Auffällig ist die lokal begrenzte Verbreitung der meisten CC121-Isolate, so dass viele Isolate ausschließlich in einer einzigen geographisch limitierten Region gefunden wurden. So umfasst Linie A mit 13 *S. aureus*-Isolaten überwiegend Isolate aus Ozeanien, mit Ausnahme von Isolat 10-01134, das aus den USA stammte. Linie B ist neben einem deutschen (Haplotyp 104) und 3 französischen Isolaten (Haplotypen 108-109; 116) zu 75 % mit Isolaten diverser afrikanischer Länder assoziiert. In Linie C stammen mehr als die Hälfte der Isolate aus Nord-, Süd- und Mittelamerika. Linie D wird mit nur einer Ausnahme (Haplotyp 72) durch europäische Isolate repräsentiert. Eine ähnliche Assoziation besteht auch bei Linie E. Sie besteht bis auf ein französisches Isolat (Haplotyp 91) ausschließlich aus asiatischen Isolaten (Kambodscha und Thailand). Linie F wird, ähnlich wie Linie D, hauptsächlich (32 von 35) durch europäische Isolate vertreten. Lediglich zwei amerikanische (Haplotyp 47 und 48) und ein afrikanisches Isolat aus Algerien (Haplotyp 53) sind ebenfalls Linie F zuzuordnen. In den Abb. 5.3 und 5.4 fallen vor allem *S. aureus*-Isolate aus Frankreich auf, da sie im Vergleich zu den anderen Isolaten phylogenetisch gesehen eine größere Diversität bezüglich ihrer geographischen Herkunft aufweisen. Französische Isolate können in einer Vielzahl verschiedener phylogenetischer Linien wiedergefunden werden.

Um festzustellen, ob die Korrelationen zwischen geographischer Verbreitung und Phylogenie statistisch signifikant sind, wurden die Datensätze mit der Software BaTS analysiert. Diese wurde bereits in Abschnitt 4.4 auf Seite 37 erläutert. Weicht die von BaTS berechnete Null-Hypothese einer zufälligen Verteilung signifikant von der tatsächlichen Verteilung der entsprechenden Herkunftsländer über die Phylogenie mit  $p \leq 0,05$  ab, wird von einer signifikanten Assoziation zwischen der geographischen Herkunft der Isolate und der Phylogenie ausgegangen. Die entsprechenden Ergebnisse der BaTS-Analyse sind in Tabelle 5.2 aufgeführt.

Tabelle 5.2: Ergebnisse der BaTS-Analyse zur Phylogenie/Geographie Assoziation

Region	MC-Wert	p-value
Afrika	5,08	0,01
Asien	3,34	0,01
Europa	9,08	0,01
Mittelamerika	10,45	0,01
Nordamerika	2,00	0,05
Südamerika	1,17	1,00
Ozeanien	3,00	0,01
<b>AI</b> =2,56; p=0		
<b>PS</b> =25,72; p=0		



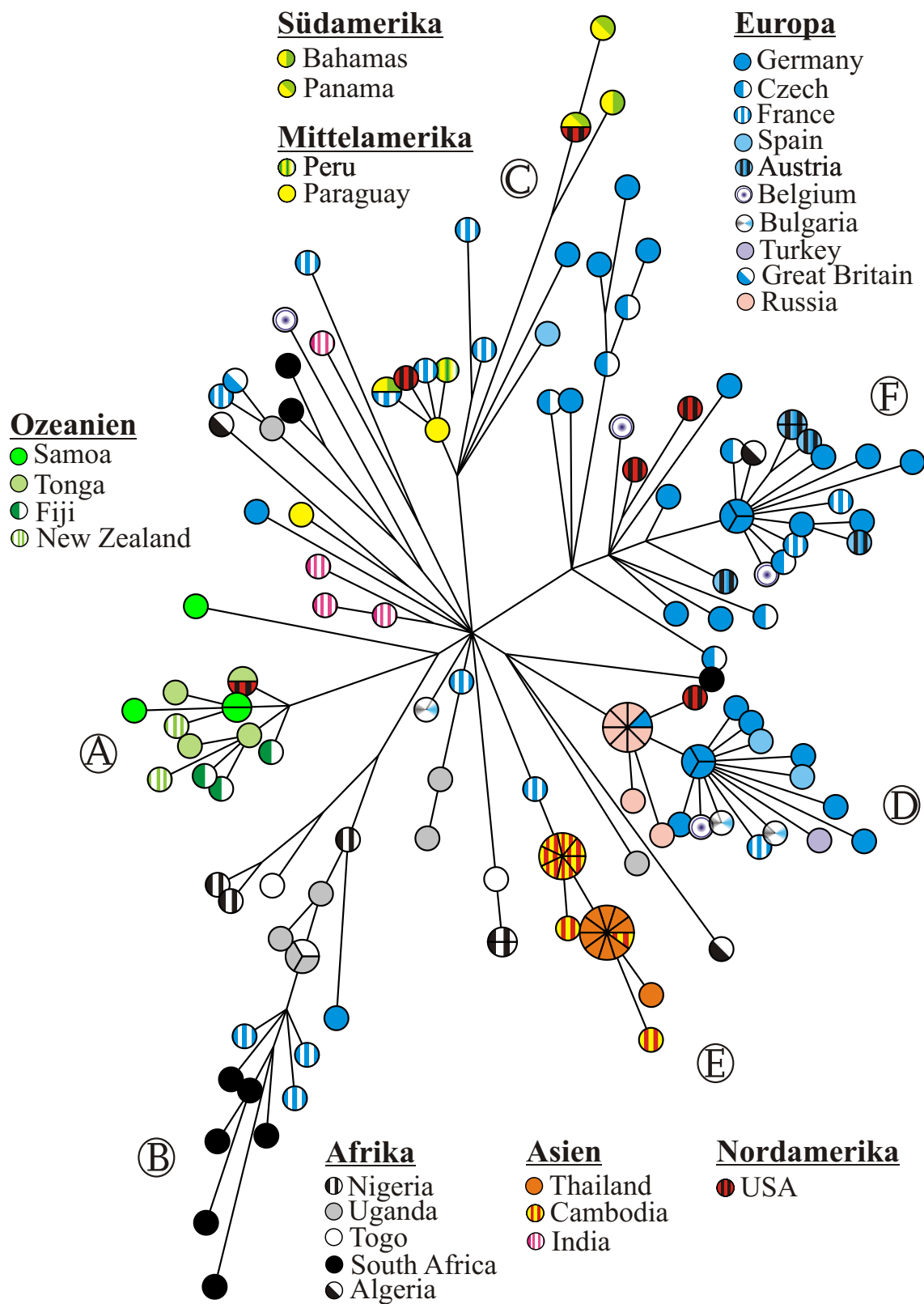


Abbildung 5.4: *Minimum-Spanning Tree*; Farbige Darstellung der geographischen Herkunft der einzelnen CC121-Isolate

### 5.3.4 Klinischer Phänotyp

Die in Abbildung 5.5 dargestellten MSTs beschreiben den Zusammenhang von klinischem Phänotyp und Phylogenie. In Abbildung 5.5-I sind für 46 *S. aureus*-Isolate die klinischen Phänotypen farbig markiert. Diese epidemiologischen Daten lagen für 30 % der in der Studie untersuchten Isolate vor. Das Fehlen entsprechender Daten für ein Großteil der Isolate beruht einerseits auf der lückenhaften epidemiologischen Dokumentation vieler Institutionen bzw. des behandelnden Personals und andererseits darauf, dass nicht alle Isolate im Zusammenhang mit einer Infektion gefunden wurden, sondern manche auch aufgrund einfacher Kontrollstudien identifiziert wurden. In Abbildung 5.5-II wurden alle klinischen Krankheitsbilder aus Abb. 5.5-I als tiefgehende oder oberflächliche Infektionen zusammengefasst.

Oberflächliche Infektionen umfassen dabei ausschließlich Infektionen, die die oberen Schichten der Haut (Epidermis) betreffen. Dazu gehören Impetigo contagiosa, *Staphylococcal scaled skin syndrome* (SSSS) und die Konjunktivitis (siehe Abb.5.5-II und 5.7).

Dagegen sind tiefgehende Infektionen nicht auf die Epidermis beschränkt und führen meist zu invasiven und pyogenen *S. aureus*-Infektionen in der tiefen Dermis und Subkutis. Zu dieser Gruppe gehören beispielsweise die akute bakterielle Infektion der Skelettmuskulatur (Pyomyositis), nekrotische Pneumonie, Abszesse und Furunkel ebenfalls zusammengefasst (siehe Abb.5.5-II und 5.7). Isolate, bei denen der klinische Phänotyp keine eindeutige Zuordnung in tiefgehende oder oberflächliche Infektionen erlaubte, wurden nicht berücksichtigt. Letzteres traf für elf der 46 Isolate zu. Hiervon waren sechs Isolate durch eine Wundinfektion charakterisiert worden und fünf mit nicht näher definierten *skin and soft tissue infections* (SSTI).

Beim direkten Vergleich der zwei Infektionstypen sowohl in Abbildung 5.5-II als auch in Abb.5.7 wird deutlich, dass alle oberflächlichen Hautinfektionen der Linie F zuzuordnen sind. Isolate, die für tiefgehende Infektionen verantwortlich sind, verteilen sich dagegen über mehrere phylogenetische Linien mit Ausnahme von Linie F. Demnach bilden die oberflächlichen Hautinfektionen in Linie F eine monophyletische Gruppe (Klade), die von den phylogenetischen Sublinien tiefergehender Infektionen differenziert werden kann.

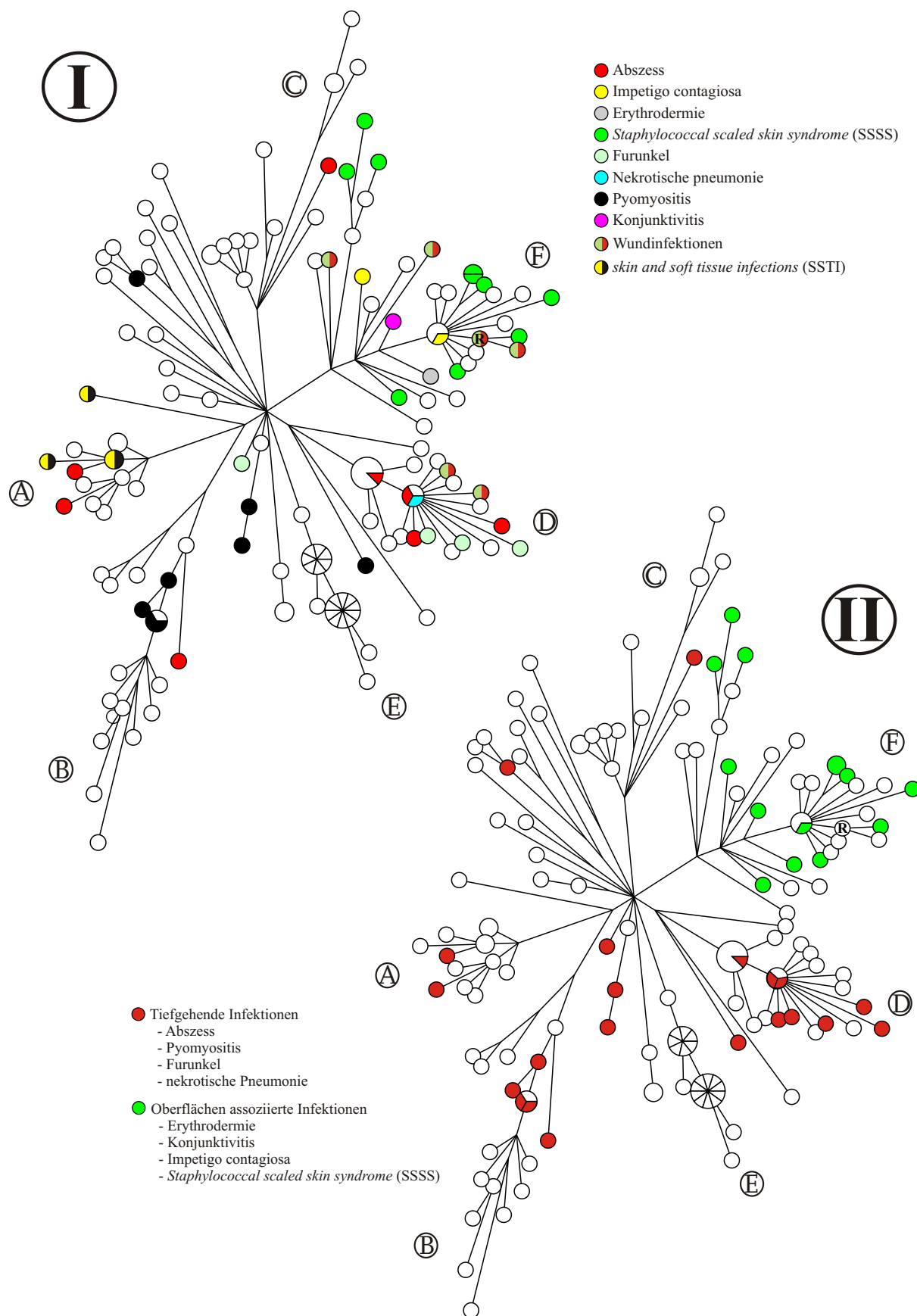


Abbildung 5.5: *Minimum-Spanning Tree*; Bild I zeigt klinische Phänotypen entsprechender Isolate; Bild II fasst tiefgehende (rot) und oberflächliche Infektionen (grün) zusammen.

### 5.3.5 Verteilung der Toxingene *eta*, *etb* und *lukS/F-PV*

Um einen Eindruck von der phylogenetischen Verteilung der getesteten Toxingene (*eta*, *etb* und *lukS/F-PV*) in den 154 untersuchten CC121-Isolaten zu erhalten, wurden diese im MST und ML-Baum farbig markiert (Abb.5.6 und Abb.5.7). Bei der Untersuchung der Toxingene fiel auf, dass in keinem der 154 untersuchten Isolate Exfoliatine und PVL zusammen vorkamen. Für Sieben Isolate konnte ausschließlich ETA nachgewiesen werden (blau markiert) und für drei ausschließlich ETB (gelb markiert) (siehe Abb.5.6 und Abb.5.7). Nicht markiert wurden Isolate, die keine der drei Virulenzdeterminanten trugen (neun Isolate). Aus den Abbildungen 5.6 und 5.7 wird deutlich, dass *lukS/F-PV* positive Isolate in allen phylogenetischen Sublinien bis auf Linie F vorkamen. Die Verteilung der Exfoliativen Toxine A und B war dagegen begrenzter. Alle positiv getesteten Isolate bis auf zwei (Haplotyp 27 und 31) entstammten der Linie F.

Diese Ergebnisse korrelieren mit denen zum klinischen Phänotyp (Abb.5.5-II und 5.7). Isolate, die nachweislich für oberflächliche Hautinfektionen verantwortlich waren, entstammten ausschließlich der Linie F, während Isolate, die tiefgehende Infektionen auslösen, sich divers über viele phylogenetische Linien verteilten, in Linie F aber nicht vorkamen (Abb.5.5-II und 5.7). Damit sind oberflächliche Hautinfektionen mit der Präsenz von Exfoliativen Toxingenen assoziiert, die in Linie F vorkommen (siehe Abb. 5.5-II und 5.6). Dagegen gehen tiefgehende Infektionen mit *lukS/F-PV* einher, das in mehreren phylogenetischen Linien gefunden werden kann (siehe Abb. 5.5-II und 5.6). Diese phylogenetische Differenzierung verstärkt den Hinweis auf eine getrennte Evolution zwei klinisch unterscheidbarer Subpopulationen.

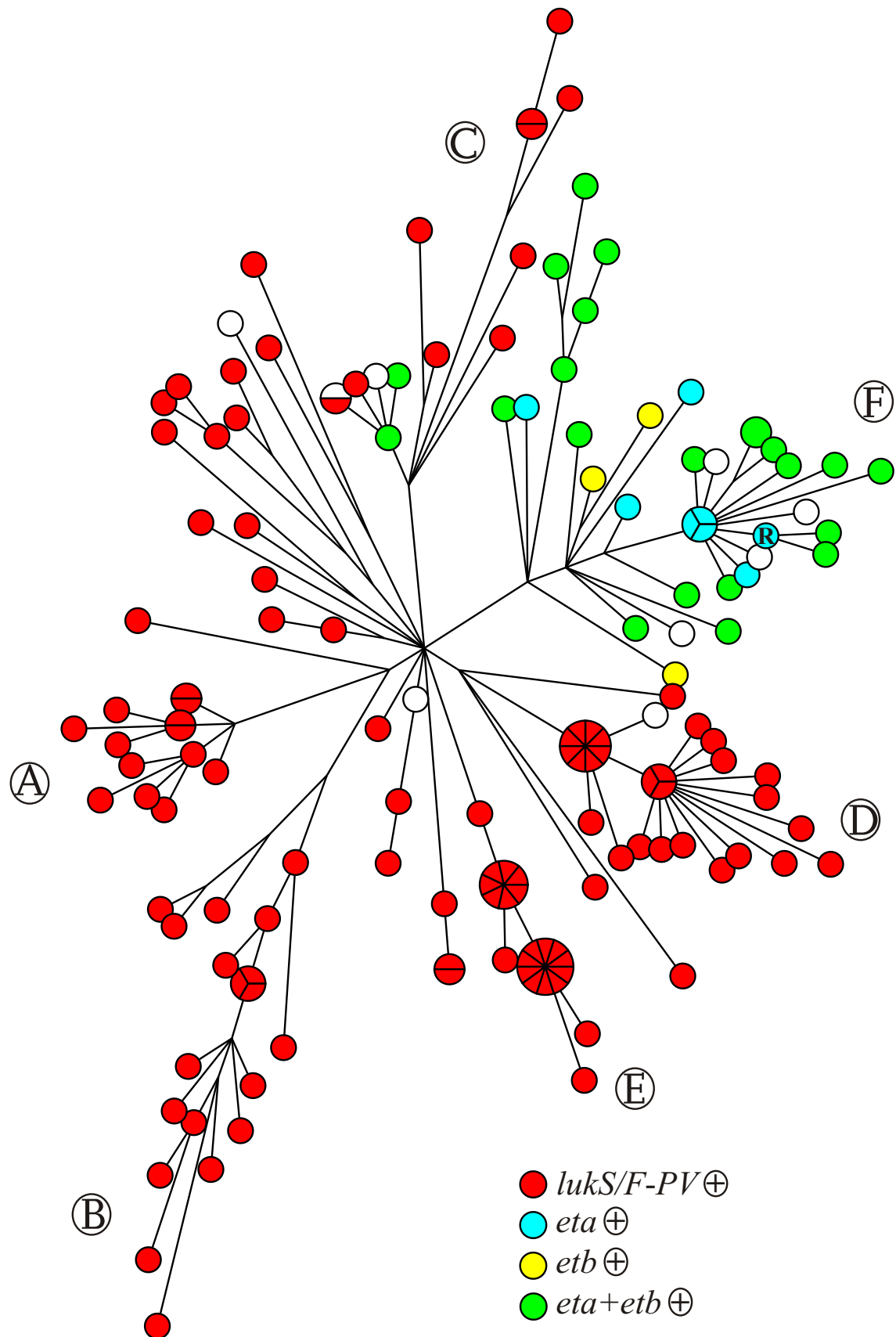


Abbildung 5.6: Die Toxine ETA, ETB und PVL sind entsprechend ihrer phylogenetischer Lokalisation auf einen *Minimum-Spanning Tree* gemappt und mit entsprechenden Farben markiert.

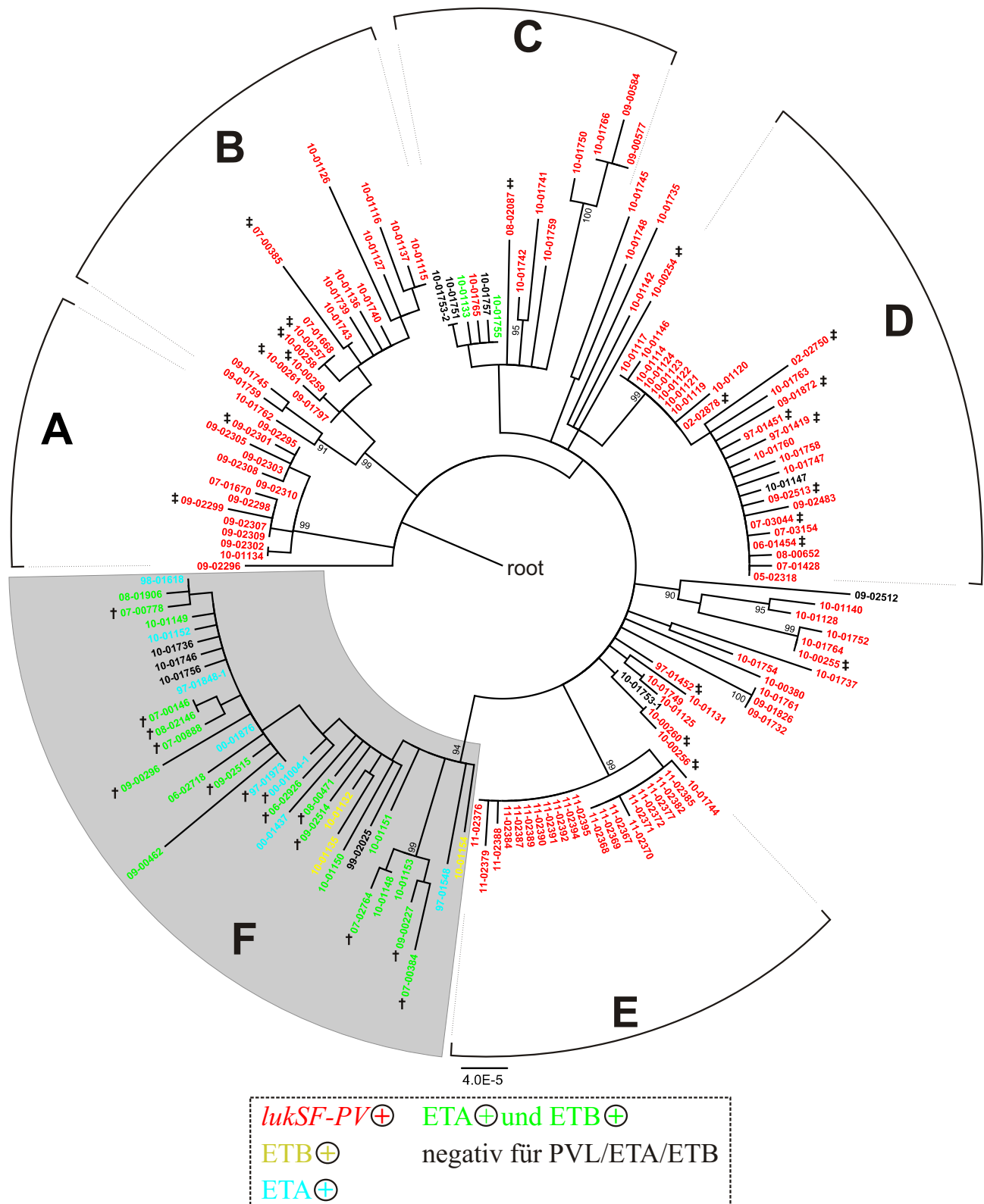


Abbildung 5.7: *Maximum-Likelihood*-Baum zeigt die Verteilung der Virulenzgene *eta*, *etb* und *lukS/F-PV*; berechnet aus konkateniertem Sequenzalignment von 154 *S. aureus*-Isolaten (46,811 bp) mittels Seaview; Virulenzgene wurden farbig unterlegt und nur Bootstrap-Werte über 90 % an den entsprechenden Kanten im Baum eingefügt. Ein Kreuz „†“ charakterisiert Isolate, bei denen eine oberflächliche Hautinfektion (Konjunktivitis, Erythrodermie, Impetigo contagiosa oder *SSSS*) identifiziert werden konnte. Während tiefegehende Infektionen wie Pyomyositis, nekrotische Pneumonie, Abszesse und Furunkel mit einem Lothringer Kreuz „‡“ markiert wurden.

### 5.3.6 Assoziation zwischen Phylogenie und Toxingenen sowie klinischem Phänotyp

Der in Abbildung 5.7 dargestellte ML-Baum zeigt den phylogenetischen Zusammenhang der drei Virulenzdeterminanten ETA, ETB und PVL mit dem klinischen Phänotyp, der für 35 Isolate validiert werden konnte. Die Isolate wurden entsprechend der detektierten Toxine farbig unterlegt und abhängig von der Art der verursachten Infektionen (oberflächlich oder tiefgehend) mit entsprechenden Symbolen gekennzeichnet (Abb.5.7). Dabei konnte in einer spezifischen phylogenetischen Klade – in Klade F – ein Zusammenhang zwischen oberflächlich assoziierten Hautinfektionen und der Existenz von Exfoliativen Toxinen beobachtet werden. Dagegen waren alle klinisch validierten tiefgehenden Infektionen mit den Virulenzgenen *lukS/F-PV* assoziiert, wiesen aber im Gegensatz zu den Isolaten mit oberflächlich assoziierten Hautinfektionen eine deutlich größere Diversität auf, da sie in allen Kladen mit Ausnahme von Klade F vorkamen.

Um festzustellen wie statistisch signifikant die Korrelation zwischen Phylogenie und Toxingen-Gehalt und dem klinischem Phänotyp ist, wurde der Datensatz mit der Software BaTS analysiert. Die Ergebnisse dieser Analysen können den Tabellen 5.3 und 5.4 entnommen werden. Dabei wurden signifikante Zusammenhänge zwischen der Phylogenie und Isolaten mit oberflächlichen und tiefgehenden Infektionen (siehe Tab.5.3), als auch zwischen der Phylogenie und dem Vorkommen der Virulenzgene *eta*, *etb* und *lukS/F-PV* ermittelt (siehe Tab.5.4). Somit konnte nicht nur signifikant eine Assoziation zwischen der Phylogenie und dem Toxingen-Gehalt nachgewiesen werden, sondern auch zwischen Phylogenie und klinischem Phänotyp. Demnach sind Isolate, ursächlich für oberflächlichen Hautinfektionen, nicht nur statistisch signifikant mit exfoliativen Toxinen assoziiert, sondern clustern auch signifikant in einer spezifischen phylogenetischen Linie, Klade F. Während Isolate, verantwortlich für tiefgehende Infektionen, assoziiert mit *lukS/F-PV* sehr divers über die gesamte Phylogenie verteilt sind, mit Ausnahme von Klade F.

Tabelle 5.3: BaTS-Werte für die Assoziation zwischen Phylogenie und klinischem Phänotyp

Infektionstyp	MC-Wert	p-value
oberflächlich	13,00	0,01
tiefgehend	13,83	0,01
AI=2,69E-06; p=0		
PS=1,00; p=0		

Tabelle 5.4: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Virulenzgen-Gehalt

Virulenzgen	MC-Wert	p-value
<i>lukS/F-PV</i>	35,98	0,01
<i>eta/etb</i>	5,07	0,01
AI=1,80; p=0		
PS=11,84; p=0		

### 5.3.7 Untersuchung zur Sequenzvariation der *lukS/F-PV* Gene

Die bereits von O'Hara *et al.* 2008 und Brown *et al.* 2012 publizierten Sequenzpolymorphismen in den Virulenzgenen *lukS-PV* und *lukF-PV* wurden im Rahmen dieser Arbeit ebenfalls untersucht [13, 113]. Bei der Analyse der 112 *lukS/F-PV* positiv getesteten Isolate konnten insgesamt zehn verschiedene Sequenzpolymorphismen in den beiden *lukS/F-PV* Genen identifiziert werden (siehe Tabelle 5.5).

Tabelle 5.5: Sequenzpolymorphismen in der genetischen Determinante für PVL (*lukS/F-PV*)

PVL-Genotyp	Anzahl Isolate	SNP-Position im <i>lukS/F-PV</i> -Gen									
		282 <sup>a</sup>	286 <sup>b</sup>	335 <sup>a</sup>	367 <sup>a</sup>	527 <sup>b</sup>	663 <sup>b</sup>	1186 <sup>a</sup>	1396 <sup>b</sup>	1694 <sup>a</sup>	1729 <sup>b</sup>
R1	1	A	C	C	C	G	T	C	A	G	G
R2	9	A	C	C	C	G	T	C	A	G	<b>A</b>
Neu-I	1	A	C	C	C	G	<b>G</b>	C	A	G	<b>A</b>
H1	1	A	C	C	C	<b>A</b>	<b>G</b>	C	<b>G</b>	G	<b>A</b>
H2	60	A	C	C	C	<b>A</b>	<b>G</b>	C	A	G	<b>A</b>
Neu-II	5	A	C	C	C	<b>A</b>	<b>G</b>	<b>T</b>	A	G	<b>A</b>
Neu-III	4	A	C	C	<b>T</b>	<b>A</b>	<b>G</b>	C	A	G	<b>A</b>
Neu-IV	1	<b>T</b>	C	C	C	<b>A</b>	<b>G</b>	C	A	G	<b>A</b>
H3	27	A	C	C	C	<b>A</b>	T	C	A	G	<b>A</b>
H3*	1	A	<b>T</b>	C	C	<b>A</b>	T	C	A	G	<b>A</b>
Neu-V	1	A	C	C	<b>T</b>	<b>A</b>	<b>G</b>	C	A	<b>A</b>	<b>A</b>
Neu-VI	1	A	C	<b>T</b>	C	<b>A</b>	<b>G</b>	C	A	G	<b>A</b>

<sup>a</sup>Neu-identifizierte Sequenzpolymorphismen

<sup>b</sup>bereits publiziert [113, 13]

Die in der Tabelle fett hervorgehobenen Nukleotide repräsentieren Sequenzpolymorphismen zum Referenzstamm USA300 (Akzessionsnummer: CP000730), der dem PVL-Genotyp R1 entspricht und grau-unterlegte SNP-Positionen kennzeichnen nicht-synonyme Mutationen. Die Auswahl des Referenzstammes USA300 geht auf eine vorausgegangene Publikation zurück [113]. O'Hara *et al.* 2008 klassifizierten die H und R Variante als zwei Grundvarianten des PVL-Genotyps [113]. Die beiden Varianten, unterscheiden sich nur durch eine nicht-synonyme Mutation in Position 527. Dabei kommt es zum Austausch von einem Guanin durch ein Adenin und damit zu einer Aminosäurenänderung von Arginin (R) zu Histidin (H). Die weitere Subklassifizierung der zwei Grundvarianten erfolgt anhand zusätzlicher Sequenzpolymorphismen, die sowohl nicht-synonyme als auch synonyme Mutationen umfassen [113, 13].

Zusätzlich zu den bereits von O'Hara *et al.* 2008 publizierten Sequenzpolymorphismen, konnten im Rahmen der vorliegenden Analyse noch fünf neue Sequenzpolymorphismen in den *lukS/F-PV*-Genen detektiert werden (Tab.5.5). Zehn Isolate konnten der R-Variante zugeordnet werden und 89 Isolate der H-Variante. Bei 13 Isolaten war keine eindeutige Zuordnung möglich, da in diesen Isolaten neue Sequenzpolymorphismen identifiziert wurden (Tab.5.5).

Abbildung 5.8 zeigt die phylogenetische Verteilung der unterschiedlichen PVL-Geno-



typen innerhalb der CC121-Population in einem ML-Baum. Die verschiedenen identifizierten PVL-Genotypen sind unterschiedlich farbig markiert. Neu identifizierte Sequenzvarianten sind durch entsprechende Synonyme gekennzeichnet (Neu-I bis Neu-VI). Es fällt auf, dass sich Isolate des PVL-Genotyps H3 vorwiegend – mit Ausnahme von vier Isolaten (10-01741, 10-01735, 10-01142 und 97-01452) – in Klade D befinden, während der Genotyp H2 über mehrere phylogenetische Linien verteilt ist. Die R-Varianten R1 und R2 sind ausschließlich in Klade E lokalisiert. Des Weiteren wird aus den Daten ersichtlich, dass PVL-Prophagen mindestens 14 Mal voneinander unabhängig akquiriert wurden (Abb.5.8) und dass einzelne PVL-Genotypen in unterschiedlichen Taxa unabhängig voneinander entstanden sind (z.B. H3 und Neu-III).

Von den ST121-MRSA-Isolaten (Abb.5.8) konnte ausschließlich 11-02369 als R2-Variante charakterisiert werden, während die vier anderen ST121-MRSA-Isolate dem PVL-Genotyp H2 zugeordnet werden müssen.

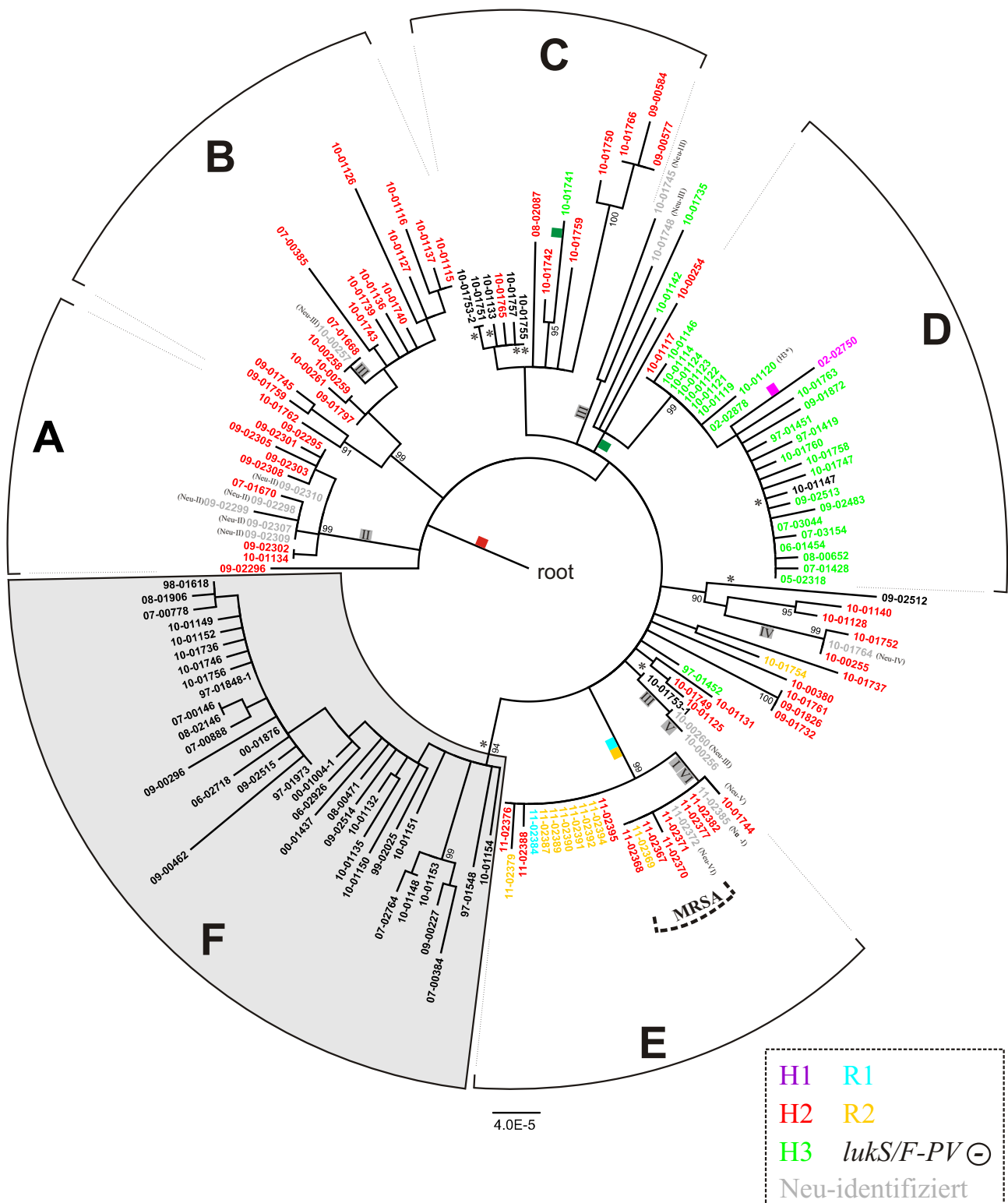


Abbildung 5.8: Maximum-Likelihood-Baum beschreibt Verbreitung der unterschiedlichen PVL-Genotypen; berechnet aus konkateniertem Sequenzalignment von 154 *S. aureus*-Isolaten (46,811 bp) mittels Seaview; PVL-Genotypen wurden farbig unterlegt und nur Bootstrap-Werte über 90 % an den entsprechenden Kanten im Baum eingefügt; Die Farbmarkierungen an den Ästen sind äquivalent zu den Farbmarkierungen der PVL-Genotypen und deuten die möglichen Aufnahmeevents der unterschiedlichen PVL-Prophagen an; Verluste von *lukS/F-PV* sind durch ein \* gekennzeichnet.

### 5.3.8 Evaluation der *spa*-Typen

In der untersuchten CC121-Population konnten 38 verschiedene *spa*-Typen identifiziert werden (Tab.A.1). Die Stammsammlung wird zu 67 % durch die sechs *spa*-Typen t159, t284, t314, t435, t645 und t940 dominiert (siehe Abb.5.9). Den größten Anteil davon stellt *spa*-Typ t159 mit 50 Isolaten, gefolgt von t645 mit 17 Isolat-en, t314 mit 14 Isolaten, t284 mit acht Isolaten und t940 mit fünf Isolaten. Einige *spa*-Typen sind mit bestimmten phylogenetischen Kladen assoziiert. Insbesondere Isolate mit dem *spa*-Typ t314 kommen ausschließlich in Klade B, Isolate mit dem *spa*-Typ t940 in Klade C und Isolate mit dem *spa*-Typ t435 in Klade D vor. Abbildung 5.10 zeigt die *spa*-Typen, die sich in der vorhandenen Phylogenie durch Homoplasien auszeichnen. Von den 38 verschiedenen *spa*-Typen trifft dies auf die *spa*-Typen t170, t272, t284, t645 und t2155 zu, da sie sich über mehrere phylogenetische Linien erstrecken und vermutlich mehrere Male unabhängig voneinander entstanden sind.

Auch in diesem Fall wurde die Software BaTS verwendet, um eine mögliche Korrelation von *spa*-Typen und SNP-basierter Phylogenie zu analysieren.

Tabelle 5.6: Ergebnisse der BaTS-Analyse zur Phylogenie/*spa*-Typ Assoziation

<i>spa</i> -Typ	MC-Wert	<i>p-value</i>
t645	5,01	0.01
t314	4,53	0,01
t284	1,30	1
t940	3,08	0,01
t435	1,38	1
t170	1	1
t272	1	1
t2155	1	1
<b>AI</b> =7,07; p=0		
<b>PS</b> =51,13; p=0		

Die Ergebnisse der BaTS-Analyse (siehe Tab.5.6) decken sich zum Teil mit den anfänglich aufgestellten Vermutungen über mögliche Korrelationen. So konnte keine signifikante Korrelation zwischen *spa*-Typ t435 und Klade D ermittelt werden, während für *spa*-Typ t645 eine Korrelation mit der Phylogenie berechnet werden konnte.

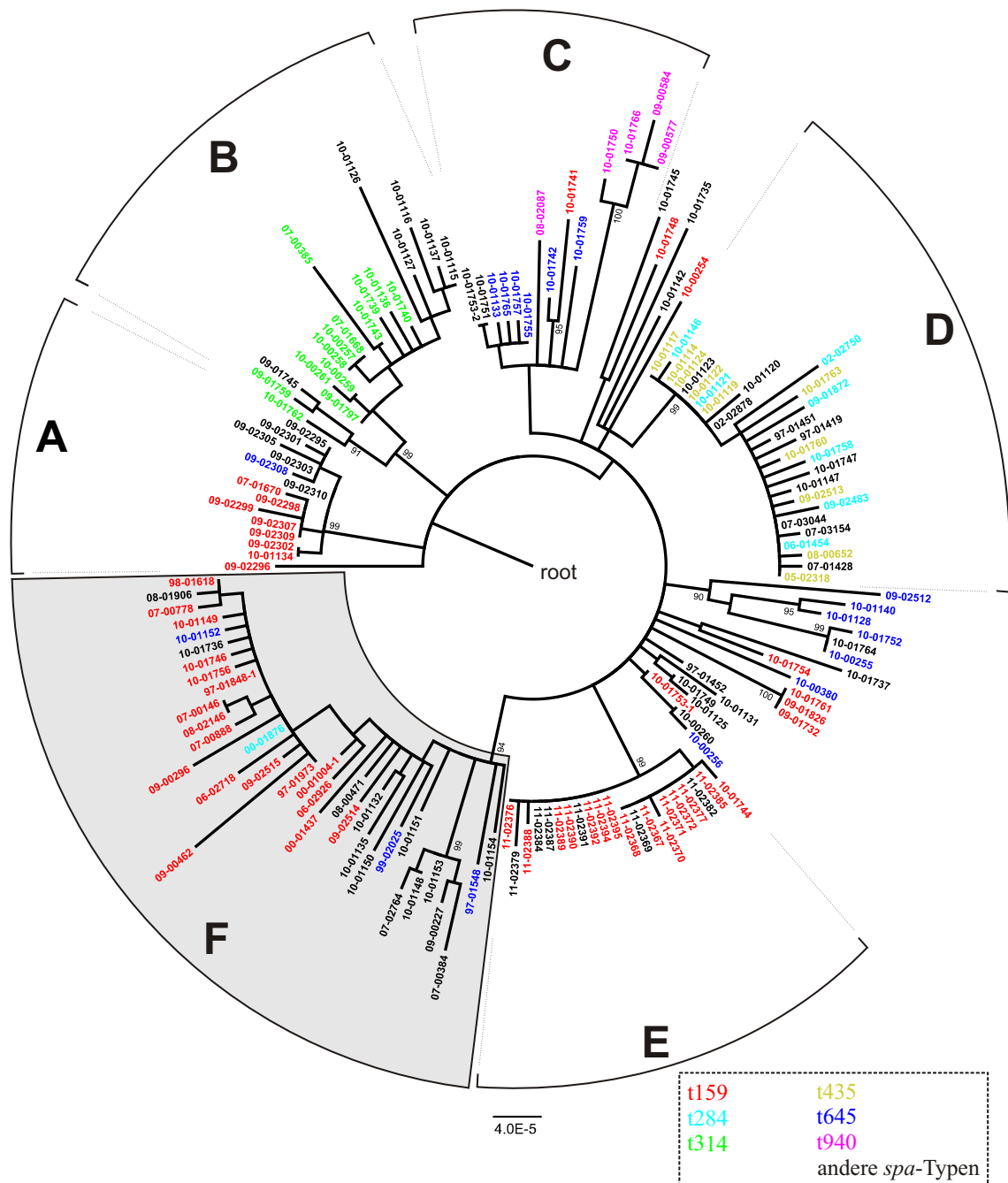


Abbildung 5.9: *Maximum-Likelihood-Baum* zeigt die Verteilung der sechs häufigsten *spa*-Typen; berechnet aus konkateniertem Sequenzalignment von 154 *S. aureus*-Isolaten (46,811 bp) mittels Seaview; *spa*-Typen wurden farbig markiert und nur Bootstrap-Werte über 90 % an den entsprechenden Kanten im Baum eingefügt.

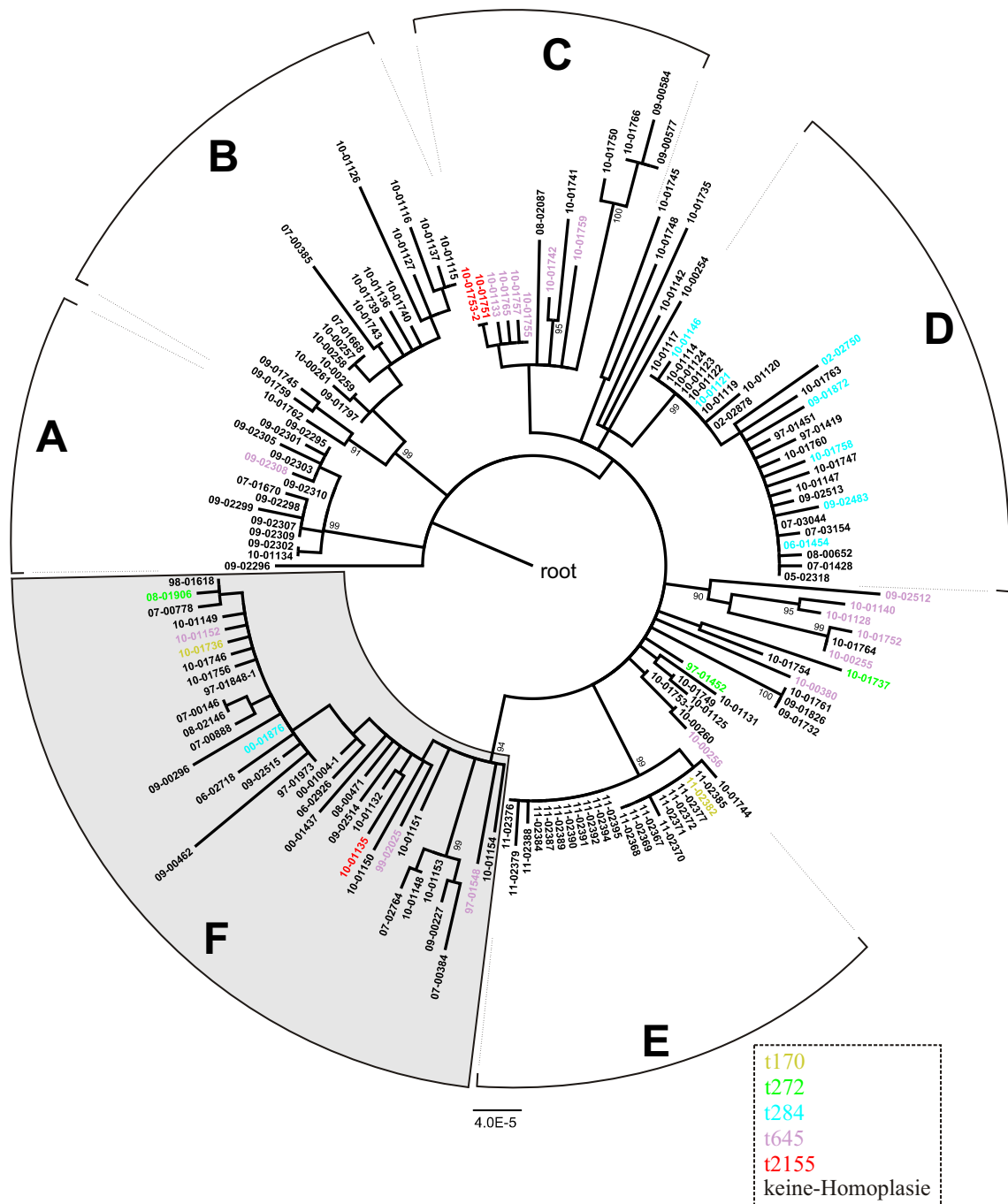


Abbildung 5.10: *Maximum-Likelihood*-Baum zeigt für fünf *spa*-Typen alle gefundenen Homoplasien; berechnet aus konkateniertem Sequenzalignment von 154 *S. aureus*-Isolaten (46,811 bp) mittels Seaview; *spa*-Typen wurden farbig markiert und nur Bootstrap-Werte über 90 % an den entsprechenden Kanten im Baum eingefügt.

## 5.4 Sequenzalignment-basierte Statistiken

### 5.4.1 Analysen zur molekularen Uhr mittels BEAST

Die Modellierung der Evolutionsraten und die Datierung des letzten gemeinsamen Vorfahren (engl. “*time to most recent common ancestor*”, TMRCA) erfolgte im Rahmen der phylogenetischen Analyse der untersuchten CC121-Population mit der Software BEAST. In Tabelle 5.7 sind die Ergebnisse für verschiedene “*clock models*”, und unterschiedliche “*tree priors*”, aufgeführt. Zwischen den ermittelten Evolutionsraten gibt es keine signifikanten Unterschiede. Des Weiteren sind auch die Entstehungszeitpunkte des letzten gemeinsamen Vorfahren ausgehend vom jüngsten Isolat (2009) aufgeführt.

Tabelle 5.7: Ergebnisse der BEAST-Analyse für die Berechnung der Evolutionsraten und des TMRCA

<i>Tree Prior</i>	<i>Clock Model</i>	<b>Evolutionsrate</b> (95 % KI) Austausche / Nukleotid*Jahr	<b>TMRCA</b> (95 % KI)
Bayesian Skyline	relaxed (lognormal)	$1,7 \cdot 10^{-6}$ ( $4,0 \cdot 10^{-7}$ - $3,1 \cdot 10^{-6}$ )	141 (59-235) Jahre
Bayesian Skyline	strict	$1,6 \cdot 10^{-6}$ ( $5,0 \cdot 10^{-7}$ - $2,9 \cdot 10^{-6}$ )	137 (60-242) Jahre
Constant Size	strict	$2,0 \cdot 10^{-6}$ ( $1,2 \cdot 10^{-6}$ - $2,8 \cdot 10^{-6}$ )	125 (84-182) Jahre

### 5.4.2 Ermittlung von $dN/dS$

Unter Verwendung der Software DnaSP v5.10.01 war es möglich, in den kodierenden Regionen synonyme und nicht-synonyme Basensubstitutionen zu unterscheiden. Von den 304 detektierten SNPs konnten 93 als synonym und 211 als nicht-synonyme Basensubstitutionen identifiziert werden (Tab.A.2). Des Weiteren ist es möglich, auf der Basis dieser Daten den Einfluss der Selektion zu beleuchten. Dieser lässt sich aus dem Verhältnis ( $dN/dS$ ) der Anzahl der nicht-synonymen Substitutionen (dN) zu der Anzahl der synonymen Mutationen (dS) ableiten. Für den  $dN/dS$ -Wert ergab sich im Mittel ein Wert von 0,475.

### 5.4.3 Analyse zum Homoplasie-Index

Um ein Maß für Homoplasie innerhalb des untersuchten Datensatzes zu erhalten, wurde die Software PAUP\* v4.0b10 benutzt. Mit PAUP\* konnte das vorher konzipierte Sequenzalignment der untersuchten CC121-Population im NEXUS-Format eingelesen und der Homoplasie-Index (HI) bestimmt werden. Der HI-Wert von 0,0297 ähnelt den Werten vorangegangener Studien zu ST5 und ST22 [110]. Durch die Analyse konnten drei der 304 identifizierten SNPs (Tab.A.2) als homoplastisch identifiziert werden. Diese Befunde stehen im Konflikt zur ermittelten Phylogenie. Das bedeutet, dass die Ähnlichkeit zwischen verschiedenen Taxa nicht mehr ausschließlich

über die Vererbung gemäß eines gegebenen Baumes erklärt werden kann. Folgende SNPs konnten als homoplastisch bestätigt werden: au208-2, au258-2, au277-1.

#### 5.4.4 Nukleotid-Diversität ( $\pi$ )

Die Nukleotid-Diversität ( $\pi$ ) wurde mithilfe der Software DnaSP auf Basis des Sequenzalignments berechnet. Sie entspricht der Wahrscheinlichkeit, mit der zwei zufällig ausgewählte homologe Nukleotide unterschiedlich sind. Die Nukleotid-Diversität ist ein Maß für die Variabilität eines bestimmten Bereichs des Genoms und beschreibt darüber hinaus die durchschnittliche Anzahl der Nukleotidunterschiede einer Sequenz zwischen zwei zufällig ausgesuchten Proben einer Bezugspopulation [102]. Der ermittelte  $\pi$ -Wert  $0,00026 \pm 0,00001$  ist um den Faktor drei größer als bei ST5 [111].

## 5.5 Genomanalysen

Analysen vollständiger Genomsequenzen bieten einen vertieften Einblick in die Evolution der Population. Zudem können zusätzlich Informationen über Virulenzeigenschaften und über die Akquirierung sogenannter Pathogenitätsinseln, Prophagen und anderer mobilen Elementen gewonnen werden. Die phylogenetische Positionierung einzelner Isolate in den spezifischen Kladen wurde im Rahmen der vorliegenden Arbeit bisher auf der Basis von Mutationen in 1,7 % des Genoms vorgenommen. Um den gesamten genomischen Kontext zu beleuchten, waren Untersuchungen vollständiger Genome notwendig. Dafür wurden zehn Isolate ausgewählt und mit der NGS-Technologie 454 vollständig sequenziert. Die Tabelle 5.8 enthält eine Übersicht der sequenzierten Isolate mit den dazugehörigen Statistiken der *Read*-Auswertungen und des *Assembly*'s. Um zu gewährleisten, dass die Auswahl der Isolate repräsentativ für die gesamte Populationsstruktur von CC121 war, wurden alle vorkommenden monophyletischen Kladen (A - F) abgedeckt (Abbildung 5.11). Neben dem phylogenetischen Aspekt sollte auch die MRSA- und MSSA-Evolution in ST121 untersucht werden. Aus diesem Grund wurden die drei Isolate 11-02371, 11-02369 und 11-02376 aus Klade E mit in die Ganzgenomanalysen aufgenommen. Die Gründe für die Auswahl dieser drei Isolate werden im Abschnitt 5.6.6 auf Seite 79 im Detail erörtert.

Tabelle 5.8: Statistiken zur 454-Sequenzierung und des Assemblierens

Isolate	Anzahl <i>Reads</i>	Sequenzierte Basen [bp]	Basen in <i>Con-</i> <i>tigs</i> [bp]	Anzahl <i>Contigs</i>	Anzahl Ns	<i>Coverage</i>
09-02298	111.165	59.241.301	2.799.811	68	7.320	21
10-00380	199.999	104.485.792	2.774.845	62	11.275	37
11-02369	83.392	36.895.275	2.859.777	129	185.317	13
11-02371	327.905	180.932.539	2.866.945	64	2.919	65
11-02376	77.932	35.391.021	2.814.173	156	243.238	13
98-01618	147.836	54.701.540	2.732.815	66	8.197	20
10-01133	152.876	60.987.325	2.867.429	103	5.823	26
10-00259	180.473	61.887.660	2.765.831	93	6.097	24
09-00227	169.098	56.147.449	2.802.022	92	8.555	22
09-02483	267.879	153.093.600	2.762.071	201	336.442	55



Abbildung 5.11: *Maximum-Likelihood*-Baum, beschreibt das Isolate Sampling für die Genom-Analysen; berechnet aus konkateniertem Sequenzalignment von 154 *S. aureus*-Isolaten (46,811 bp) mittels Seaview; rot markierte Isolate repräsentieren Isolate, die für die Genomsequenzierung mittels NGS-Technologie ausgewählt wurden.

### 5.5.1 Phylogenie

Für die Rekonstruktion der Phylogenie wurde ein SNP-Alignment verwendet, das auf 1828 variablen qualitätsgeprüften Positionen des Kerngenoms basiert, an denen sich die zehn untersuchten Isolate voneinander unterscheiden. Das SNP-Alignment schließt nur variable Positionen des Kerngenoms ein, da im Vorfeld das akzessorische Genom mit allen mobilen genetischen Elementen (Prophagen, SaPIs, Plasmide, Transposons, Insertionselemente und der *SCCmec*-Kassette), *Repeat*-Regionen und Bereiche, die nicht in allen Genomen vorkommen, ausgeschlossen wurden. Um den phylogenetischen Baum gewurzelt darzustellen, wurde auch bei diesen Analysen das Kerngenom von N315 mit in die Analysen aufgenommen. Die Abbildung 5.12 zeigt zwei aus der ML-Methode und der Bayes'schen Analyse abgeleitete Stammbäume. Die Position der Wurzel ist in beiden Stammbäumen identisch mit derjenigen, die bei der Untersuchung der 46,811 bp (1,7 % des Genoms) ermittelt wurde. Beide Stammbäume weisen eine identische Topologie auf, wobei lediglich die statistische Unterstützung der Verzweigungsknoten in einigen Werten variiert. Die Isolate 09-02298 aus Klade A, 10-00259 aus Klade B und 10-01133 aus Klade C bilden eine gemeinsame Klade, was sich von dem Ergebnis der SNP-basierten Untersuchung unterscheidet.

Wie bereits in Abschnitt 5.3.2 beschrieben wurde, ist auch in Abbildung 5.12 das Monophylum für die Isolate 98-01618 und 09-00227 aus Klade F sichtbar. Eine andere monophyletische Gruppe wird von den drei thailändischen Isolaten 11-02369, 11-02371 und 11-02376 gebildet. Die beiden MRSA-Isolate unterscheiden sich durch 75 SNPs voneinander und sind damit im Vergleich zum vorliegenden MSSA-Isolat 11-02376 mit 163 SNPs deutlich näher miteinander verwandt. Diese Befunde decken sich mit den bereits in Abschnitt 5.3.2 beschriebenen Ergebnissen. Unabhängig von der Methode (Ganzgenomanalyse, Analyse von 1,7 % des Genoms) sind die populationsgenetische Struktur und phylogenetische Beziehung der selektierten Isolate identisch.

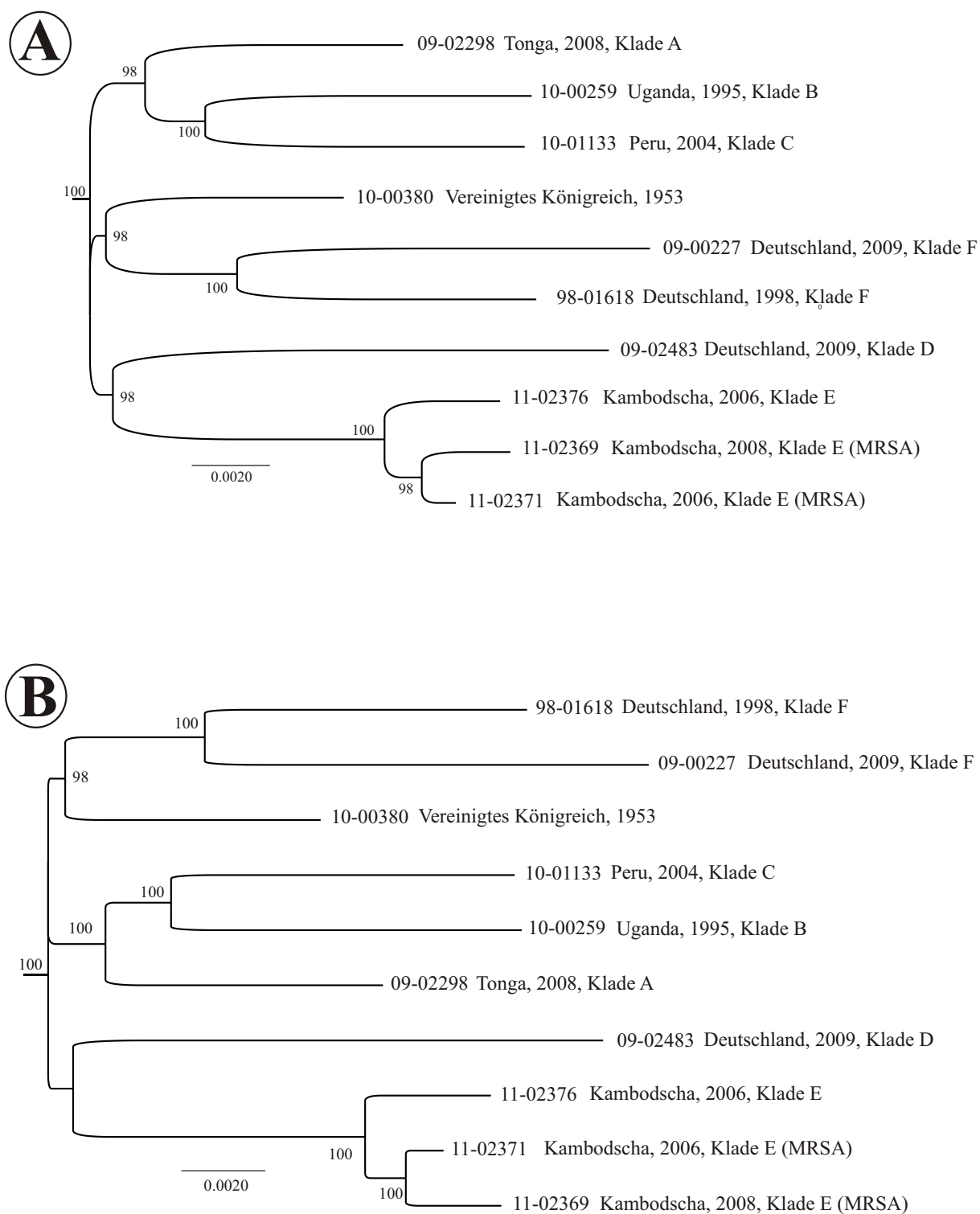


Abbildung 5.12: Phylogenie der zehn vollständig sequenzierten ST121-Genome, basierend auf 1828 SNPs. A: *Maximum-Likelihood*-Baum; B: Bayes'scher Baum. Bootstrap-Werte über 90 % sind an den entsprechenden Kanten angegeben.

## 5.6 Genomvergleiche

Die anhand der repräsentativen Stichprobe (zehn CC121-Isolate) durchgeführten Genomanalysen und die Identifizierung von mobilen genetischen Elementen erfolgte mithilfe der Software Kodon. Dabei konnte der überwiegende Teil der Diversität der sequenzierten Isolate auf das akzessorische Genom mit einer großen Anzahl von mobilen genetischen Elementen zurückgeführt werden (siehe Tab.5.9).

Tabelle 5.9: Detektierte Sequenzpolymorphismen im *lukS/F-PV*-Gen

MGE <sup>a</sup>	Integrationsstelle [ORF in N315]	Serotyp <sup>b</sup>	Integrase- Gruppe	Größe [kb]	Charakteristische Gene <sup>c</sup>	98-01618(F)	09-00227(F)	10-01133(C)	10-00259(B)	09-02483(D)	09-02298(A)	10-00380	11-02369(E)	11-02371(E)	11-02376(E)
SaPI1	<i>downstream</i> von <i>groEL</i> , zwischen ORF [SA1815 und SA1836]			~ 27	<i>tst</i>							x			
	s.o.			~ 29	keine Virulenzgene	x									
	s.o.			~ 18	<i>fusB</i>		x			x	x				
SaPI2	<i>upstream</i> des Gens <i>guaA</i> , [SA0376]			~ 29	<i>set</i> [9-10], <i>lpl</i> [3]	x	x	x	x	x	x	x	x	x	x
SaPI3	<i>downstream</i> eines t-RNA Clusters, zwischen ORF [SA1621-1649]			~ 22	<i>lukDE</i> , <i>egc</i> [5], <i>sp</i> [4], <i>ear</i> , <i>hsdM2</i>	x	x	x	x	x	x	x	x	x	x
SaPI4	<i>upstream</i> des Gens <i>rpsR</i> , [SA0354]			~ 17	<i>seb</i>			x	x				x	x	x
Sa1 Prophage	<i>attB</i> in [SA0778]	Bc-Phage	Sa1	~ 42	<i>eta</i>	x	x	x							
	s.o.	Bc-Phage	Sa1	~ 42	keine Virulenzgene					x				x	
Sa2 Prophage	integriert in [SA1320]	A-phage	Sa2	~ 46	<i>lukSF-PV</i>				x	x	x	x	x	x	x
Sa3 Prophage	in Gen <i>hbl</i> , [SA1811]	Fa-Phage	Sa3	~ 44	<i>sak</i> , <i>chp</i> , <i>scn</i>			x		x					
	s.o.	Fb-Phage	Sa3	~ 36	<i>sak</i> , <i>scn</i>		x		x	x	x	x	x	x	x
Sa5 Prophage	<i>attB</i> in [SA1693]	Ba-Phage	Sa5	~ 33	keine Virulenzgene			x							
	s.o.	Bc-Phage	Sa5	~ 33	keine Virulenzgene								x		x
ΦSaCC121	integriert zw. [SA1150 und SA1154]	unbekannt	unbekannt	~ 41	keine Virulenzgene	x	x	x	x	x	x	x	x	x	x
ETB-Plasmid				~ 38	<i>eth</i> , <i>edin-C</i> , <i>cadDX</i>		x	x							
SCCmecV	<i>orfX</i> [SA0023]			~ 42	<i>mecA</i> , <i>ccrC1allel8</i> , <i>ccrC1allel10</i>								x	x	
<b>Resistom<sup>c</sup></b>							<i>blaZ</i> <i>fusB</i> <i>ermC</i>	<i>blaZ</i> <i>tetK</i>	<i>blaZ</i> <i>tetK</i>	<i>blaZ</i> <i>ermC</i> <i>mecA</i>	<i>blaZ</i> <i>ermC</i>	<i>blaZ</i> <i>ermC</i> <i>mecA</i>	<i>blaZ</i> <i>ermC</i> <i>mecA</i>	<i>blaZ</i> <i>tetK</i>	<i>blaZ</i> <i>tetK</i>

<sup>a</sup>Graustufen entsprechen verschiedenen Varianten einer Gruppe, die sich aufgrund vom Gengehalt unterscheiden lassen<sup>b</sup>Nach R. Pantucek *et al.*, 2004 [120]<sup>c</sup>Angegeben sind Gene, die in Virulenz oder Antibiotika-Resistenz involviert sind; Abkürzungen: *ccrC*, Rekombinase-Gen C (*cassette chromosome recombinase* C); *lukDE*, Komponenten des Leukozidin DE Toxins; *lukS/F-PV*, Komponenten des Panton-Valentine Leukozidin S und F; *mecA*, Methicillin-Resistenzprotein 2a; *set*, Staphylokokken Exotoxine; *spl*, Staphylokokken Serinprotease; *tst*, Toxin-1 des Toxisch-Schock-Syndroms; *egc*, Enterotoxin-Cluster; *eta* und *eth*, Exfoliatives Toxin A und B; *fusB*, Fusidinsäure-Resistenz Protein FusB; *sak*, Staphylokinase; *seb*, Staphylokokken Enterotoxin B; *scn*, Staphylokokken Komplement Inhibitor SCIN; *chp*, Chemotaxis-Inhibitor protein CHIPS; *hbl*,  $\beta$ -Hämolsin; *ear*, vermutlich  $\beta$ -Laktamase Protein; *blaZ*,  $\beta$ -Laktamase Gen für Penicillinresistenz; *ermC*, Erythromycin-Resistenz; *cadDX*, Cadmium Resistenz Operon; *edin-C*, Epidermaler Zelldifferenzierung Inhibitor C; *tetK*, Tetracyclin-Resistenz

### 5.6.1 Pathogenitätsinseln

Pathogenitätsinseln (PAIs) sind weit verbreitet und wurden für gram-positive Bakterien erstmals von Lindsay *et al.* in *S. aureus* beschrieben [86]. Im Rahmen der vorliegenden Analyse konnten in den zehn sequenzierten CC121-Isolaten insgesamt 31 Pathogenitätsinseln identifiziert werden (siehe Tab.5.9). Diese können aufgrund ihrer Integrationsstelle im *S. aureus*-Genom und auf Basis der Integrase-Homologie in verschiedene Gruppen eingeteilt werden [45].

**SaPI1.** Die Pathogenitätsinsel SaPI1 integriert *downstream* des Chaperonin GroEL kodierenden *groEL*-Gens (SA1815; Locus Tag in N315) und wird in der Literatur oft mit verschiedenen Enterotoxinen (*sel* und *sec3*) und dem Toxic-shock-syndrome-Toxin 1 (*tst-1*) in Verbindung gebracht. Fünf der 31 Pathogenitätsinseln konnten in fünf der sequenzierten Isolate aufgrund ihrer Integrationsstelle im Genom als Pathogenitätsinsel SaPI1 identifiziert werden (Tab.5.9). Der Vergleich mit bereits publizierten SaPI1-Varianten zeigte eine große interne genetische Diversität.

Bei der Pathogenitätsinsel in Isolat 10-00380 konnte im Gegensatz zu den übrigen vier als herausstechendes Merkmal das Exotoxin kodierende Toxigen *tst-1* identifiziert werden (Tab.5.9). In Isolat 09-00227 konnte mit einer fast 99 %igen Sequenzidentität die bereits von O'Neill *et al.* 2007 publizierte Pathogenitätsinsel mit der Bezeichnung *SaRI<sub>fusB</sub>* detektiert werden (Tab.5.9) [116]. Diese wurde speziell in Zusammenhang mit *epidemic European fusidic acid-resistant impetigo clone*-(EEFIC)-Isolaten des ST123 beschrieben. Sie kodiert unter anderem für das Resistenzgen *fusB* und vermittelt so den betreffenden Isolaten die Resistenz gegenüber der tetracyclischen Triterpensäure Fusidinsäure. In den Isolaten 98-01618, 09-02483 und 09-02298 lassen sich keine bekannten Virulenz- und Resistenzdeterminanten innerhalb der Pathogenitätsinsel identifizieren. Darüber hinaus können generell nur vereinzelt strukturelle Ähnlichkeiten zwischen den SaPI1 der Isolate festgestellt werden (Tab.5.9).

**SaPI2.** Die Pathogenitätsinsel SaPI2 integriert *upstream* des Gens *guaA*, dass für das Enzym Guanosin-monophosphat-(GMP)-Synthase kodiert. Charakteristisch für die Pathogenitätsinsel SaPI2 ist, dass sie Gene für ein Exotoxin-Gen-Cluster (*set*), ein Lipoprotein-Gen-Cluster (*lpl*) und für die zwei Determinanten *hsdS* und *hsdM* des Restriktionsmodifikationssystems (RM-Systems) beinhaltet [6]. Die Pathogenitätsinsel SaPI2 konnte in allen analysierten Isolaten detektiert werden (Tab.5.9). In allen Isolaten konnten neun bis zehn *set*-Homologe sowie je drei *lpl*-Homologe identifiziert werden. Die Gene *hsdS* und *hsdM* des RM-System waren zu 99 % sequenzidentisch mit den bereits von Waldron and Lindsay 2006 beschriebenen Genen, *Sau1hsdS1* und *hsdM2* [152]. Insgesamt ist SaPI2 stark konserviert und es gibt nur wenige Bereiche mit Indels und SNPs.

**SaPI3.** Wie auch SaPI2 konnte die Pathogenitätsinsel SaPI3 für alle sequenzierten Genome nachgewiesen werden. SaPI3 integriert spezifisch *downstream* eines *tRNA*-Clusters und kodiert für die Gene eines Serin-Protease-Clusters (*spl*), eines Enterotoxin-Clusters (*egc*) und für die Gene *lukD/E*, die für zwei Komponenten des Leukozidin-Toxins LukD-LukE kodieren. Genau wie SaPI2 beinhaltet auch SaPI3 die beiden Gene *hsdS* und *hsdM* des RM-Systems [6]. In allen sequenzierten Isolaten konnten sowohl vier paraloge Gene des Serin-Protease-Clusters als auch die Gene für die beiden Komponenten des Leukozidin-Toxins LukD-LukE detektiert werden (Tab.5.9). Des Weiteren war es möglich, alle fünf Gene des Enterotoxin-Clusters, die für homologe Enterotoxine kodieren (*seo*, *sem*, *sei*, *sen*, und *seg*), zu identifizieren. Auffallend war, dass SaPI3 in allen sequenzierten Isolaten nur für das Gen *hsdM* kodiert und nicht, wie in der Literatur beschrieben, auch für das Gen *hsdS* [6]. Das identifizierte *hsdM* Gen in SaPI3 stimmte, genau wie bei SaPI2 beschrieben, zu 99 % mit dem von Waldron and Lindsay 2006 beschriebenen *hsdM2* Gen überein [152]. Die Gen-Determinante *hsdS* des RM-Systems lag in allen sequenzierten Isolaten in der Pathogenitätsinsel SaPI3 trunkiert vor. SaPI3 ist innerhalb der CC121-Linie stark konserviert und unterscheidet sich im Vergleich zu der bereits beschriebenen SaPI3 im Genom von N315 nur in wenigen Sequenzabschnitten.

**SaPI4.** In den sechs Isolaten 09-02298, 10-01133, 10-00259, 11-02369, 11-02371 und 11-02376 konnte mit SaPI4 eine weitere Gruppe der Pathogenitätsinseln detektiert werden (Tab.5.9). SaPI4 integriert *downstream* des Gens *rpsR*, das für ein ribosomales Protein der 30S-Untereinheit kodiert [58]. Neben einigen wenigen Genabschnitten, die Homologien zur SaPI1 von N315 zeigen, stimmt der Rest der Sequenz der identifizierten SaPI4 kaum mit bereits publizierten SaPIs überein. Die Pathogenitätsinsel SaPI4 lag in allen sechs Isolaten stark konserviert vor und enthielt neben einigen Genen, die für hypothetische Proteine oder für die spezifische Integrase kodieren, auch das Virulenzgen *seb*. Dieses Virulenzgen kodiert für das Staphylokokken-Enterotoxin B (SEB) und gehört zu den sogenannten Superantigenen [66].

## 5.6.2 Prophagen

Prophagen können bezüglich ihrer genomischen Zusammensetzung sehr variabel sein und einen erheblichen Einfluss auf die Diversität einer Population nehmen. In den zehn sequenzierten Genomen konnten insgesamt 34 Prophagen mit einer Größe von ca. 33 kb bis 46 kb identifiziert werden (Tab.5.9). Jedes der untersuchten Isolate enthielt zwei bis maximal vier Prophagen in seinem Genom. Die analysierten Prophagen konnten alle – bis auf einen, der im späteren Ergebnissteil noch weiter vertieft wird – der Virusfamilie *Siphoviridae* zugeordnet werden, die Bakteriophagen der Ordnung *Caudovirales* umfasst. Die Virusfamilie zeichnet sich durch ikosaedrische Kapside und nicht-kontraktile Schwanzteile aus. Diese konnten nach dem System von Goerke *et al.* (2009) anhand ihrer Integrase-Gene in vier verschiedene int-Gruppen (Sa1, Sa2, Sa3 und Sa5) eingeteilt werden (Tab.5.9) [47]. Darüber hinaus war es durch die von Pantucek *et al.* (2004) publizierten Primer, die *in silico* verwendet wurden, möglich die verschiedenen Prophagen in serologische Gruppen einzuteilen [120].

**Sa1.** Die spezifische Integrationsstelle des Sa1-Prophagen (*attachment site* oder *recombination site*) befindet sich im *sufB*-Gen (ORF-Bezeichnung in N315: SA0778), das Teil des Operon-Komplexes *sufBCD* ist. SUF-Transporter sind an der Assemblierung von Eisen-Schwefel-Clustern beteiligt. Fünf der 34 Prophagen konnten der int-Gruppe Sa1 zugeordnet werden und in fünf der sequenzierten Isolate identifiziert werden (Tab.5.9). Die Diversität innerhalb der Prophagen reicht von Regionen mit großen strukturellen Ähnlichkeiten über solche mit einzelnen SNPs bis hin zu Regionen, die kaum Ähnlichkeiten aufweisen. Im Gegensatz zu den beiden Isolaten 11-02371 und 09-02483 konnten die drei Isolate 98-01618, 09-00227 und 10-01133 mit dem Toxingen *eta*, das für das exfoliative Toxin A kodiert, assoziiert werden (Tab.5.9). Abbildung 5.13 zeigt das Alignment der fünf Sa1-Prophagen in den betreffenden Isolaten, sowie des bereits von Yamaguchi *et al.* (2000) publizierten ETA-Phagen  $\Phi$ ETA (Akzessionsnummer:AP001553) [161]. Das Alignment umfasst eine Länge von 41,868 bp und zeigt, dass alle Sa1-Prophagen dem ETA-Phagen  $\Phi$ ETA in großen Teilen des Genoms ähneln. Besonders konserviert sind die Gene, die für die DNA-Verpackung und für die Morphogenese der Phagenpartikel verantwortlich sind (siehe Abbildung 5.13).



## Sa1

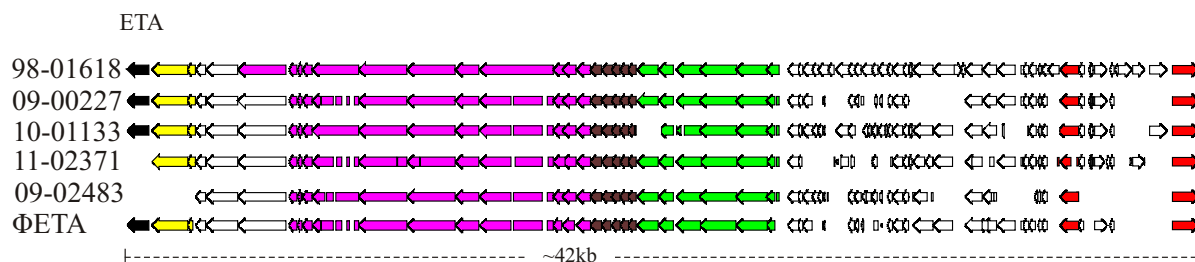


Abbildung 5.13: Genetische Organisation der Sa1-Prophagen im Alignment; Dargestellt sind alle fünf Isolate, in denen ein Sa1-Prophage identifiziert werden konnte, im Vergleich mit dem bereits publizierten Prophagen ΦETA. Bestimmte Prophagen-Module konnten bioinformatisch zugeordnet werden und sind farbig markiert: für die Lyse verantwortliche Gene, gelb; Strukturgenen für Kopf, Schwanz, Schwanzplatte und Schwanzfasern, magenta, braun und grün; für die Lysogenie verantwortliche Gene, rot; Virulenzdeterminanten *lukSF-PV*, schwarz.

**Sa2.** Prophagen der Integrase-Gruppe Sa2 konnten in sieben der zehn Isolate identifiziert werden (Tab.5.9). Die spezifische Integrationsstelle dieser Prophagen befindet sich in einem Gen mit unbekannter Funktion (ORF-Bezeichnung in N315: SA1320) [101]. Abbildung 5.14 zeigt das Alignment aller sieben Sa2-Phagen der jeweiligen Isolate mit dem bereits durch Baba *et al.* (2002) publizierten Prophagen ΦSa2mw, der als PVL-Phage im *S. aureus*-Genom MW2 (Akzessionsnummer: BA000033) identifiziert wurde und strukturelle Ähnlichkeiten zu den detektierten Prophagen zeigt [6]. Generell ist die Sequenzähnlichkeit bei den verschiedenen Prophagen unterschiedlich ausgeprägt. Diese reicht von großen strukturellen Ähnlichkeiten über einzelne SNPs (von 6 bis > 750 SNPs) bis hin zu Regionen, die kaum noch eine Ähnlichkeit aufweisen. Auffallend war die hohe Sequenzübereinstimmung der beiden Prophagen aus den beiden asiatischen MRSA Isolaten 11-02371 und 11-02376, bei denen sich die Prophagen nur um sechs SNPs unterschieden. Alle identifizierten Sa2-Phagen kodieren die Gene *lukF-PV* und *lukS-PV* für das 2-Komponenten-Toxin Pantone-Valentine-Leukozidin.

## Sa2

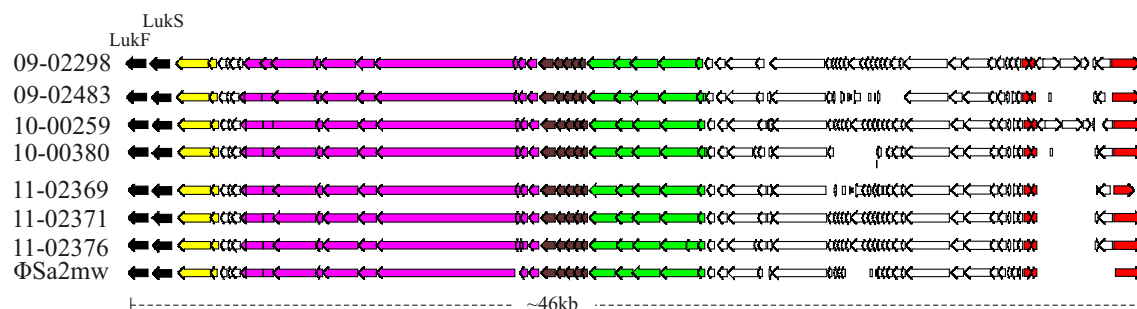


Abbildung 5.14: Genetische Organisation des Sa2-Prophagen im Alignment; Dargestellt sind alle sieben Isolate, in denen ein Sa2 Prophage identifiziert werden konnte, im Vergleich mit dem bereits im Genom MW2 identifizierten Prophagen ΦSa2mw. Bestimmte Prophagen-Module konnten bioinformatisch zugeordnet werden und sind farbig markiert: für die Lyse verantwortliche Gene, gelb; Strukturgenen für Kopf, Schwanz, Schwanzplatte und Schwanzfasern, magenta, braun und grün; für die Lysogenie verantwortliche Gene, rot; Virulenzdeterminanten *lukSF-PV*, schwarz.

**Sa3.** Bei der Integration des temperenten Phagen der Integrase-Gruppe Sa3 kommt es zunächst zu einer negativen lysogenen Konversion, da die spezifische Integra-

tionsstelle innerhalb des Gens *hly* liegt, das für das Toxin  $\beta$ -Hämolysin kodiert (ORF-Bezeichnung in N315; SA1811), und die Integration zum Ausfall des Genproduktes führt [20]. Die Analysen ergaben neun verschiedene Prophagen der Integrase-Gruppe Sa3 in neun Isolaten (Tab.5.9). Alle neun Sa3-Prophagen zeigen neben der negativen Hly-Konversion auch eine positive Konversion für verschiedene Virulenzfaktoren (Sak, CHIPS und SCIN). Einer der neun Prophagen in Isolat 10-01133 kodiert Gene für die Staphylokinase (*sak*), für das Antiphagozytose-Protein SCIN (*scn*) und das Chemotaxis-inhibierende Protein CHIPS (Tab.5.9). Die restlichen acht beinhalten nur die beiden Gene *sak* und *scn* (Tab.5.9).

Die Prophagen zeigen auch hier wiederum eine hohe Diversität. Dazu gehören einzelne SNPs, zusätzliche Virulenzfaktoren unterschiedlicher Größe (möglicherweise auch Hinweis auf nicht identifizierte Bereiche, da die Prophagen nicht immer komplett auf einem *Contig* lagen), aber auch die Einteilung in verschiedene Serotypen (Tab.5.9).

**Sa5.** Die Integrationsstelle von Prophagen der Integrase-Gruppe Sa5 liegt in einem Gen, dessen Funktion bisher unbekannt ist (ORF-Bezeichnung in N315; SA1693). In drei Isolaten 10-01133, 11-02369 und 11-02376 konnte jeweils ein Prophage der Integrase-Gruppe Sa5 identifiziert werden (Tab.5.9). Diese wiesen keine spezifischen Virulenzdeterminanten auf und konnten über die verschiedenen Serogruppen Ba und Bc unterschieden werden.

**Prophage mit unbekannter Integrase-Gruppe.** In allen zehn vollständig sequenzierten Genomen konnte ein bisher noch nicht in der Literatur beschriebener Prophage mit einer Größe von ca. 41 kb identifiziert werden (Tab.5.9). Dieser wurde im Rahmen der Arbeit als  $\Phi$ SaCC121 klassifiziert. Die Integrationsstelle von  $\Phi$ SaCC121 liegt in einem Genbereich, dessen Funktion bisher nicht bekannt ist (ORF-Bezeichnungen in N315: SA1150-SA1154).

$\Phi$ SaCC121 war in allen untersuchten Stämmen stark konserviert und unterschied sich in den einzelnen Isolaten lediglich durch wenige Sequenzunterschiede (6 bis 35 SNPs). Die Eingruppierung des Prophagen nach dem System von Goerke *et al.* 2008 in eine bestimmte Integrase-Gruppe gelang nicht, da zwar auf Proteinebene Sequenzähnlichkeiten mit Integrase-ähnlichen Enzymen gefunden werden konnten, aber auf Nukleotidebene signifikante Sequenzähnlichkeiten fehlten. Insgesamt konnten 55 offene Leseraster im  $\Phi$ SaCC121 Prophagen identifiziert werden. Abbildung 5.15 zeigt eine graphische Darstellung von  $\Phi$ SaCC121 mit den dazugehörigen ORFs und deren Einteilung in die verschiedenen Genmodule des Prophagen. Mit Hilfe von BLAST (*Basic Local Alignment Search Tool*) und PHAST (*PHAge Search Tool*) wurden die Sequenzen mit bereits in einer Datenbank vorhandenen Sequenzen verglichen und nach paralogen und orthologen Genen gesucht. Dadurch war es möglich, einigen ORFs vermeintliche Funktion zuzuordnen und eine grobe Modulation der Phagenstruktur zuzulassen. Eine Aussage über einen möglichen Einfluss dieses Prophagen auf den Phänotyp der Staphylokokken kann nicht getroffen werden.

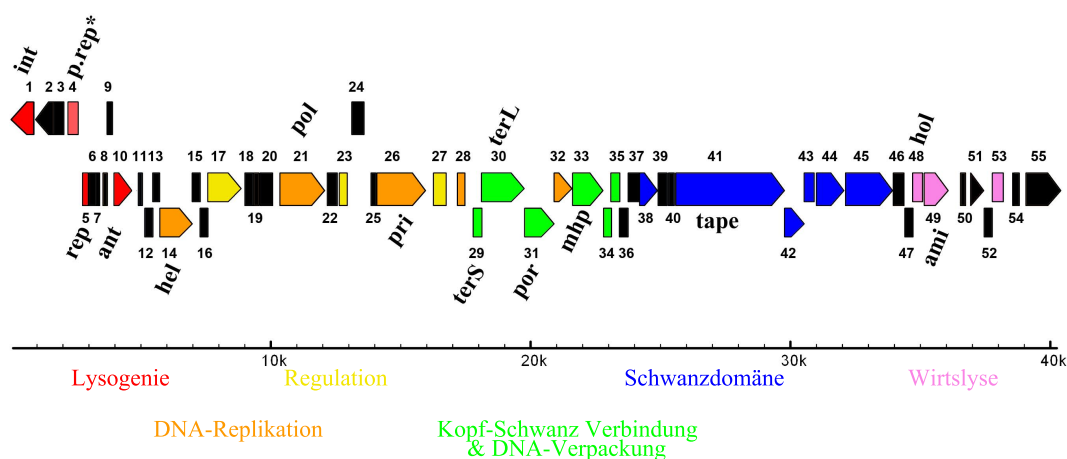


Abbildung 5.15: PHAST-Output: Genetische Organisation des  $\Phi$ SaCC121-Prophagen. Die verschiedenen bioinformatisch ermittelten Phagenmodule sind entsprechend farbig markiert. Lysogener Zyklus, rot; DNA-Replikation, orange; Regulation, gelb; DNA-Verpackung und Kopfproteine, grün; Schwanzproteine, blau; Lytischer Zyklus, magenta. Einzelne Gene sind entsprechend annotiert: int, Integrase; rep, Repressor; ant, Antirepressor; hel, Helikase; pol, DNA-Polymerase; pri, Primase; terS/terL, kleine und große Untereinheit der Terminase; por, Portal Protein; mhp, wichtiges Kopfprotein; tape, TMP (*Tape measure protein*); hol, Holin; ami, Amidase.

### 5.6.3 Plasmide

In den Isolaten 09-00227 und 10-01133 wurde im Rahmen des Genomprojekts ein Plasmid entdeckt, das die Virulenzfaktoren *etb*, *edin-c* sowie die Resistenzdeterminante *cadD* und dessen Regulatorgen *cadX* kodiert. Das Virulenzgen *etb* kodiert für das exfoliative Toxin B (ETB), *edin-c* stammt aus der EDIN-Familie (*Epidermal Cell Differentiation Inhibitor*) und *cadD* vermittelt die Resistenz gegenüber Kadmium. Das Plasmid hat eine Größe von etwa 38 kb und zeigt große Homologien zu dem erstmals in *S. aureus* TY4 (Akzessionsnummer:AP003088) beschriebenen ETB-Plasmid pETB [162]. Des Weiteren war es möglich, Plasmid kodierte Resistenzen in den Isolaten 10-01133, 11-02376, 09-00227, 09-02298, 11-02369 und 11-02371 aufzuspüren. In den Isolaten 10-01133 und 11-02376 konnte das Gen *tetK* detektiert werden, dass die Resistenz gegenüber Tetrazyklin vermittelt und für ein Efflux-Pumpen-System kodiert. Dieses System sorgt dafür, dass aufgenommene Tetrazyklinmoleküle wieder aus der Zelle hinausbefördert werden [51]. In den vier Isolaten 09-00227, 09-02298, 11-02369 und 11-02371 konnte das Resistenzgen *ermC* identifiziert werden, dass für die Resistenzentwicklung gegen Makrolide (insbesondere Erythromycin) verantwortlich ist. Über eine kovalente Bindung an die Peptidyltransferase wird der Mechanismus der Translokation blockiert und letztlich die Proteinbiosynthese inhibiert [75]. Leider kann die Anzahl und Größe der verschiedenen Plasmide nicht exakt bestimmt werden, da die beobachteten Plasmidsequenzen, Virulenzdeterminanten und Resistenzgene über einer Vielzahl von *Contigs* verteilt waren, sodass eine genaue Zuordnung nur schwer erfolgen konnte. Da eine Übertragung von Plasmid-assoziierten Resistenzgenen bereits in der Literatur mehrfach erläutert wurde, ist die Anwesenheit von Plasmiden bei derartigen Analysen anzunehmen.

### 5.6.4 Andere Virulenzdeterminanten

Bei den vergleichenden Genomanalysen wurden neben den mobilen genetischen Elementen, wie Prophagen, Plasmiden, SCC<sub>mec</sub> und Pathogenitätsinseln, noch weitere Virulenzdeterminanten, Adhäsionsprotein-kodierende Gene und regulatorische Gene untersucht. In den Tabellen 5.10, 5.11 und 5.12 sind die Ergebnisse zusammengefasst.

Tabelle 5.10: Geprüfte Adhäsions- und Oberflächenprotein codierende Gene

	<i>fnbA</i>	<i>fnbB</i>	<i>clfA</i>	<i>clfB</i>	<i>sdrC</i>	<i>sdrD</i>	<i>sdrH</i>	<i>coa</i>	<i>cna</i>	<i>isdA</i>	<i>isdB</i>	<i>isdC</i>	<i>isdD</i>	<i>isdE</i>	<i>ebps</i>	<i>ecb</i>	<i>bbp</i>	<i>sasG</i>	<i>icaA</i>	<i>icaB</i>	<i>icaC</i>	<i>icaD</i>	<i>sbi</i>
98-01618	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
09-00227	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
09-02483	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
10-01133	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
10-00259	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
10-00380	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
09-02298	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
11-02369	x	x	x	x	x	x	x	x	x	x		x	x	x	x		x		x	x	x	x	x
11-02371	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x
11-02376	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x

Aufgeführte Gene kodieren für Adhäsions- oder Oberflächenproteine; Abkürzungen: *fnbA/B*, Fibronectin (Fn)-bindende Proteine; *clfA/B*, Clumping Factor A und B; *sdrC/D/H*, Adhäsine (Serine-Aspartate-repeat containing proteins); *coa*, Koagulase; *cna*, Kollogen-Adhäsine; *isdA/B/C/D/E*, Eisen-regulierte oberflächen-determinierende Proteine A-E; *ebps*, Elastin-bindende Protein; *ecb*, Extrazelluläres Komplement-bindendes Protein; *bbp*, Bindet Bone-Sialoprotein; *sasG*, Oberflächenprotein G; *icaA/B/C/D*, Operon für Syntheseparaat des interzellulären Polysaccharid Adhäsins (PIA); *sbi*, Staphylokokken-Immunglobulin-bindendes Protein.

Tabelle 5.11: Übersicht über die getesteten Virulenzgene

	<i>edinA</i>	<i>edinB</i>	<i>splA</i>	<i>splB</i>	<i>splE</i>	<i>ACME</i>	<i>sea</i>	<i>sek</i>	<i>seq</i>	<i>sec</i>	<i>sel</i>	<i>sed</i>	<i>sej</i>	<i>ser</i>	<i>seh</i>	<i>sep</i>	<i>ORF CM14</i>	<i>sspA</i>	<i>sspB</i>	<i>sspP</i>
98-01618			x	x													x	x	x	x
09-00227			x	x													x	x	x	x
09-02483			x	x													x	x	x	x
10-01133			x	x													x	x	x	x
10-00259				x													x	x	x	x
10-00380			x	x													x	x	x	x
09-02298			x	x													x	x	x	x
11-02369			x	x													x	x	x	x
11-02371			x	x													x	x	x	x
11-02376			x	x													x	x	x	x

Aufgeführte Gene kodieren Virulenzgene; Abkürzungen: *splA/B/E*, Staphylokokken Serinproteasen A/B/E; *sea*, Enterotoxin A; *sek*, Enterotoxin K; *seq*, Enterotoxin Q; *sec*, Enterotoxin C; *sel*, Enterotoxin L; *sed*, Enterotoxin D; *sej*, Enterotoxin J; *ser*, Enterotoxin R; *seh*, Enterotoxin H; *sep*, Enterotoxin P; *edinA/B*, epidermal cell differentiation inhibitor A/B; *ACME*, Gencluster zum Argininabbau (arginine catabolic mobile element; (*arcA-D*)); *sspA/B/P*, Staphopain A/B/P; *ORF CM14*, Enterotoxin.

Tabelle 5.12: Übersicht über getestete Regulons

	<i>agrI</i>	<i>agrII</i>	<i>agrIII</i>	<i>agrIV</i>	<i>sarA</i>
98-01618				x	x
09-00227				x	x
09-02483				x	x
10-01133				x	x
10-00259				x	x
10-00380				x	x
09-02298				x	x
11-02369				x	x
11-02371				x	x
11-02376				x	x

Regulon-Gene: *agr*, akzessorischer Genregulator; *sar*, Staphylokokken akzessorischer Aktivator.

### 5.6.5 SCC*mec* Genkassette

Um zu analysieren, warum Isolate des ST121 erst vor Kurzem als MRSA auftraten, wurden zwei der fünf untersuchten ST121-MRSA-Isolate (11-02369 und 11-02371) aus Kambodscha mit in die genom-basierten Analysen aufgenommen. Von den bereits in der Literatur beschriebenen elf verschiedenen SCC*mec*-Kassetten (*International Working Group on the Staphylococcal Cassette Chromosome elements*, <http://www.sccmec.org>) tragen die beiden isolierten MRSA-Isolate eine SCC*mec*-V-Kassette. Durch die SCC*mec*-Sequenzanalyse konnte eine hohe Sequenzübereinstimmung mit dem bereits in Stamm UMCG-M4 (Akzessionsnummer:GQ902038) beschriebenen SCC*mec*-V Element festgestellt werden [17]. Abbildung 5.16 zeigt das Alignment der beiden identifizierten SCC*mec*-V Kassetten mit der bereits publizierten SCC*mec*-V Kassette in Stamm UMCG-M4. Aufgrund der Anordnung des Insertion-Elementes *IS431*, des *mecA* Gens, des trunkeierten *mecR1* Gens ( $\Delta$ *mecR1*) und einer weiteren Kopie von *IS431* in entgegengesetzter Richtung lässt sich der *mec*-Genkomplex der SCC*mec*-Kassette der Klasse C2 zuordnen [68]. In beiden SCC*mec*-Elementen sind je zwei Kopien des Rekombinase-Gens *ccrC* vorhanden. Rekombinase-Gene sind für den Ein- und Ausbau der Genkassette in die spezifische *orfX*-Region im Genom von *S. aureus* essentiell. Beide identifizierten *ccrC* Gene können durch Sequenzvergleiche und anhand der Vorgaben der *International Working Group on the Staphylococcal Cassette Chromosome elements* mit der folgenden Nomenklatur betitelt werden: *ccrC1allele8* und *ccrC1allele10* (siehe Abb. 5.16). Im Unterschied zur SCC*mec*-Kassette in UMCG-M4 liegt bei den beiden SCC*mec*-Elementen aus 11-02369 und 11-02371 keine Deletion von 1041 bp zwischen dem linken *IS431* Element und dem *mecA* Gen (vermutlich kodiert der Bereich für eine Hydroxymethylglutaryl-Coenzym A-Synthase (kurz: HMG-CoA-Synthase)), wie von Chlebowicz *et al.* 2010 in UMCG-M4 beschrieben, vor. Des Weiteren enthalten die SCC*mec*-V Elemente der beiden Isolate 11-02369 und 11-02371 zwischen dem *ccrC1allele10* Genkomplex und dem rechts flankierenden chromosomalen Bere-

ich ebenfalls Gene für ein Typ 1 RM-System. Auch in diesem Bereich bestehen große Sequenzähnlichkeiten zum Stamm UMCG-M4. Die Gen-Determinanten *hsdM*, *hsdS* und *hsdR* unterscheiden sich von denen im Stamm UMCG-M4 nur durch eine Punktmutation im *hsdR*-Gen. Hierbei handelt es sich um eine nicht-synonyme Mutation (an Position 28138 in UMCG-M4). Die Transversion von Adenin zu Thymin hat den Aminosäureaustausch von einem Glycin zu einem Arginin zur Folge.

### SCCmecV



Abbildung 5.16: Zeigt die genetische Organisation der SCCmec-Kassette der CC121-MRSA-Isolate und JCSC5952. Farblich markiert sind die SCCmec-typischen Elemente und das R-M-System

### 5.6.6 Analysen zur MRSA-Evolution in CC121

Wie bereits in Abschnitt 5.5 auf Seite 64 angedeutet wurde, war ein weiteres Ziel der Ganzgenomanalyse die Erforschung der MRSA- und MSSA-Evolution in ST121. In der Literatur wurden Isolate des ST121 bis vor Kurzem ausschließlich als MSSA beschrieben. Die vermutete Ursache war, dass die Akquirierung der SCCmec-Kassette auf bisher unbekannte Weise gestört oder gänzlich blockiert ist [107, 93]. Erst vor wenigen Jahren erschienen, ausschließlich aus dem asiatischen Raum, vereinzelt Publikationen, in denen ST121-Isolate auch als MRSA beschrieben wurden [16, 103, 38, 70].

Durch die Bereitstellung von 20 Isolaten aus der Region Kambodscha und Thailand wurde es möglich, auch diese Fragestellung im Rahmen der vorliegenden Arbeit zu untersuchen. Bei den 20 Isolaten handelt es sich um zehn thailändische und zehn kambodschanische Isolate, wobei sich fünf der thailändischen Isolate als MRSA bestätigen ließen (Tab.A.1). Um mögliche Unterschiede im Genom von MRSA und MSSA in ST121 zu untersuchen, wurden die drei Isolate 11-02371 (MRSA), 11-02369 (MRSA) und 11-02376 (MSSA) aus Kambodscha für die Genomanalyse ausgewählt.

**orfX-Region.** Zunächst wurde geprüft, ob es auffällige genomische Umstrukturierungen in der *orfX*-Region gibt, die eine mögliche Integration des SCCmec Elementes unterbinden oder gar unmöglich machen. SCCmec integriert im *S. aureus*-Genom spezifisch am 3'-Ende des *orfX*-Gens (SCCmec attachment site *attB<sub>SCC</sub>*), dessen Funktion nicht bekannt ist [62]. Für die Mobilität der SCCmec-Kassette in Form einer homologen Rekombination sind die Rekombinase-Gene (*ccrA*, *ccrB*, *ccrC*) verantwortlich [67]. Eine entscheidende Bedeutung besitzt bei diesem Prozess eine 15 bp lange Kernsequenz (TTA TGA TAC GCC TCT), die als direkte *Repeats* (Sequenzwiederholungen) flankierend an den Enden des SCCmec-Elementes auftreten

( $DR_{scc-R}$  und  $DR_{scc-L}$ ) und in *S. aureus*-Isolaten verschiedener Klonaler Komplexe hoch konserviert sind [62, 107, 90].

In keinem der zehn untersuchten Isolate konnte in der *orfX*-Region eine Irregularität gefunden werden. In allen Isolaten wurde die für die Akquirierung der *SCCmec*-Kassette spezifische *attB<sub>SCC</sub>*-Sequenz inklusive der 15 bp langen Kernsequenz im *orfX*-Gen identifiziert. Auch beim direkten Vergleich der vollständigen Genomsequenzen von MSSA und den beiden MRSA Isolaten wurden keine Unterschiede festgestellt, die eventuell auf eine bevorzugte oder verminderte Integration von *SCCmec*-Elementen hinweisen könnten.

**DNA-Umgebung der spezifischen *attB<sub>SCC</sub>*-Sequenz.** Noto *et al.* untersuchten, die chromosomale Umgebung der spezifischen *attB<sub>SCC</sub>*-Sequenz und sprachen dieser eine funktionale Bedeutung bei der Aufnahme von *SCCmec*-Elementen zu [107]. Daher wurde auch in der vorliegenden Analyse die DNA-Umgebung der spezifischen *attB<sub>SCC</sub>*-Sequenz bei allen zehn Genomsequenzen untersucht. Die Untersuchungen von Noto *et al.* (2007) konzentrierten sich auf den Vergleich von in der Literatur als MSSA beschriebenen *S. aureus*-Isolaten (darunter auch ein ST121 Isolat) mit *S. aureus*-Isolaten bekannter MRSA-Linien, wie ST8, ST1 und ST5 [107]. Sie konnten zeigen, dass die 15 bp lange Kernsequenz der *attB<sub>SCC</sub>*-Region aller untersuchten Isolate hoch konserviert ist, aber dennoch in der chromosomalen Umgebung zwischen bekannten MRSA-Linien und MSSA-Isolaten deutliche Unterschiede auftreten. Untereinander unterscheiden sich die MRSA-Isolate in dieser Region kaum voneinander, was zu dem Schluss führte, dass es einen möglichen Zusammenhang zwischen der Aufnahme von *SCCmec*-Elementen und der chromosomalen Umgebung von *attB<sub>SCC</sub>* gibt [107].

Das von Noto *et al.* verwendete ST121-Isolat 3298 zeigt im Vergleich mit den zehn sequenzierten Isolaten keine auffälligen Unterschiede in der chromosomalen Umgebung der *orfX*-Region, die auf eine funktionale Veränderung hinweisen könnten. Unterschiede beschränken sich auf eine Punktmutation in der 100 bp langen nicht für ein Protein kodierenden Umgebungssequenz, die in den beiden MRSA-Isolaten 11-02369 und 11-02371 und dem nahe verwandten MSSA-Isolat 11-02376 vorkamen.

Für alle zehn Isolate galt, dass die Umgebung von *orfX* im hohen Maße variabel war und offenbar Fragmente verschiedener Transposons enthielt. Sequenzunterschiede, die möglicherweise im Zusammenhang mit dem RM-System oder mit dem molekularen Mechanismus der Rekombination stehen, konnten nicht eindeutig identifiziert werden. Die Unterschiede zwischen den Isolaten waren vorwiegend auf das akzesessorische Genom begrenzt, das bei *S. aureus* ca. 25 % des gesamten Erbgutes ausmacht [84].



# Teil IV

## Diskussion

## 6. Diskussion

In der vorliegenden Arbeit wurden neben phylogenetischen Analysen, die auf 1,7 % des *S. aureus*-Genoms basierten, auch vergleichende Genomanalysen durchgeführt. Diese wurden zum einen dazu genutzt, die Mikroevolution der CC121-Population angesichts verschiedener epidemiologischer Daten zu untersuchen, zum anderen, um die genomische Diversität zwischen den verschiedenen phylogenetischen Sublinien mit Hilfe der Ganzgenom-Analysen zu erörtern.

Der erste Teil der Diskussion befasst sich mit der phylogenetischen Analyse von CC121 unter Berücksichtigung verschiedener epidemiologischer Fragestellungen. Im zweiten Teil wird auf die genomische Diversität der zehn für die 454-Sequenzierungstechnologie ausgewählten CC121-Stämmen eingegangen. Abschließend soll versucht werden, Aspekte der Entwicklungsgeschichte zu erläutern.

### 6.1 Mikroevolution von CC121

#### 6.1.1 Phylogenie von CC121

Die Analyse von 1,7 % des *S. aureus*-Genoms (47 kb) ergab unabhängig von der verwendeten Methode MST und ML-Methode übereinstimmende Phylogenien. Innerhalb der untersuchten CC121-Population konnten sowohl mit der Baum-Topologie des MST in Abbildung 5.2 auf Seite 45, als auch über den ML-Baum in Abbildung 5.3 auf Seite 47 sechs verschiedene Kladen (A-F) bestimmt werden.

Im Vergleich zu bereits untersuchten MRSA-Linien, wie ST22, ST225 und ST239 [55, 110], die ihre geographische Ausbreitung evolutionsgeschichtlich gesehen erst spät begonnen haben, spiegelt das tiefe Verzweigungsmuster der einzelnen Haplotyp-Cluster der CC121-Population ein relativ hohes Alter dieser Linie wieder (Abb. 5.2 und 5.3). Unterstützt wird diese Aussage durch die Ergebnisse, die sich aus der Modellierung der Evolutionsraten entlang der Zweige und aus der Datierung für den letzten gemeinsamen Vorfahren ergeben (siehe Tabelle 5.7). Die in der vorliegenden Arbeit ermittelten Daten ( $2,0 \cdot 10^{-6}$  Austausche / Nukleotid\*Jahr (95 % KI von  $1,2 \cdot 10^{-6} - 2,8 \cdot 10^{-6}$  Austausche / Nukleotid\*Jahr)) ähneln den Werten in bereits publizierten Arbeiten [55, 110]. Anhand dieser Werte und der beobachteten Sequenzvariationen wird das ungefähre Alter des letzten gemeinsamen Vorfahren (engl.,

*Most Recent Common Ancestor (MRCA)*) auf das Jahr 1884 (95 % KI von 1827 – 1925) geschätzt. Diese Schätzung ist vergleichsweise ungenau. Da Isolate aus der ersten Hälfte des 20. Jahrhunderts fehlten, mussten die vorhandenen Daten über einen langen Zeitraum extrapoliert werden. Dennoch belegen die vorliegenden Analysen, dass CC121 in der präantibiotischen Ära, lange vor dem Auftreten von MRSA, entstanden ist.

Da die beiden verwendeten phylogenetischen Baumkonstruktionsmethoden identische Ergebnisse für die globale Phylogenie der CC121-Population liefern, werden im Folgenden präferenziell die Ergebnisse der ML-Methode diskutiert. Einzelne Rückschlüsse werden zusätzlich anhand der Daten des MST gezogen. Bei der Auswahl der zehn *S. aureus*-Isolate, die mit Hilfe der 454-Sequenzieretechnologie untersucht wurden, wurde darauf geachtet, dass sie repräsentativ für die gegebene Populationsstruktur von CC121 sind und jede der sechs identifizierten Kladen (A-F) abdecken (Abbildung 5.11).

Anhand der Genomsequenzen war es möglich, die Phylogenie der CC121-Population mit einer weitaus höheren Auflösung zu rekonstruieren (Abb.5.12). Deutlich wurde das vor allem bei den Isolaten aus den Kladen A, B und C. Diese bilden eine gemeinsame Klade, was in den vorangegangenen SNP-Analysen nicht erkannt wurde (Abb.5.12). Interessanterweise zeigen sowohl die Ergebnisse der genomweiten Mutationsanalyse als auch die der SNP-basierten Phylogenieuntersuchungen eindeutig eine unterschiedliche Evolution von Teilpopulationen, die sich durch zwei unterschiedliche klinische Phänotypen voneinander absetzen. Bei beiden Analysen konnte gezeigt werden, dass die *lukS/F-PV*-positiven Isolate phylogenetisch divers sind, während die mit den Virulenzdeterminanten ETA und ETB assoziierten Isolate eine eigene monophyletische Gruppe bilden. Durch die genomweite Mutationsanalyse war es möglich, die phylogenetischen Beziehungen exakter aufzulösen und die divergente Evolution von zwei verschiedenen, klinisch relevanten Subpopulationen zu untermauern. Durch eine Auswahl der im Rahmen dieser Arbeit detektierten SNPs ist es möglich, diese evolutionär unterschiedlichen Populationen zuverlässig zu identifizieren (siehe Tab.5.1).

### 6.1.2 Phylogeographie

Die Phylogeographie beschreibt die Evolution bezogen auf die geographische Verteilung von Populationen. Sowohl der MST in Abbildung 5.4 als auch der ML-Baum in Abbildung 5.3 zeigen, dass ein Großteil der 154 bearbeiteten *S. aureus*-Isolate gemäß ihres geographischen Ursprungs sechs monophyletischen Gruppen (Klade A-F) einzuordnen sind. Die geographische Verteilung der einzelnen Isolate innerhalb dieser Kladen zeigt, eine deutliche Separation der Kladen. Angesichts dieser Daten lässt sich ein Modell für die Evolution der CC121-Population ableiten, das ein weltweites Vorkommen mit einer überwiegend lokalen Populationsstruktur pos-

tuliert. Die Isolate aus den sechs identifizierten monophyletischen Gruppen (Klade A-F) traten endemisch auf. Daraus muss geschlossen werden, dass eine interkontinentale Verschleppung ein eher seltenes Ereignis darstellt. Dieses Modell der Populationsstruktur stimmt mit den bisher publizierten Daten überein, in denen CC121 häufig als einer der bedeutenden MSSA-Klone in Amerika, Asien, Europa, Afrika und dem Pazifik beschrieben wird [151, 46, 27, 132, 44, 100]. Unterstützt wird die Auffassung einer regional geprägten CC121-Populationsstruktur durch Korrelationsberechnungen mit Hilfe der Software BaTS, die eine eindeutige Korrelation zwischen den Herkunftsländern und der Phylogenie der einzelnen Kladen ergab (siehe Tabelle 5.2). Alle untersuchten Isolate aus Europa, Afrika, Mittelamerika, Ozeanien und Asien bilden abhängig ihres geographischen Ursprungs Cluster (Tab.5.2). Eine Ausnahme stellen die Isolate aus Südamerika dar, diese zeigten keine signifikante Korrelation mit der Phylogenie; auch nordamerikanische Isolate wiesen nur eine schwach signifikante Korrelation mit der Phylogenie auf (Tab.5.2). Nübel *et al.* 2008 postulierten für die ST5-Population (für die MRSA-Population von ST5) ein ähnliches Modell für ein endemisches Auftreten [111]. Darüber hinaus konnte gezeigt werden, dass die Aufnahme der SCCmec-Elemente durch *S. aureus* wesentlich häufiger als zuvor vermutet und unabhängig voneinander erfolgt und die weltweite Verbreitung von sogenannten pandemischen MRSA-Klonen nicht auf alle MRSA zutrifft [111]. Interessant ist, dass vor allem Isolate aus der europäischen und nordamerikanischen Region im Vergleich zu Isolatzen aus anderen Gebieten deutlich diverser über die bestehenden Phylogenien verteilt sind (Abb.5.4 und 5.3). Teilweise sind die Ergebnisse für die europäischen Isolate möglicherweise durch den verwendeten Stichprobenrahmen der Stammsammlung zu erklären, da knapp 45 % der untersuchten CC121-Population aus Europa stammten.

Weiterhin fällt beim Betrachten der phylogenetischen Bäume in den Abbildungen 5.3 und 5.4 auf, dass französische Isolate häufig einen Zusammenhang mit afrikanischen Isolatzen aufweisen. Dies wird besonders für die vier Haplotypen 99, 108, 109 und 116 in Klade B deutlich (Abb.5.4). Als Ursache dafür kommen die engen historischen Beziehungen zwischen Frankreich und Afrika in Frage. Dieser Schluss kann aber aufgrund der fehlenden Rückschlüsse aus der Erhebung der epidemiologischen Daten nicht eindeutig bewiesen werden.

Interessanterweise bilden die als *epidemic European fusidic acid-resistant impetigo clone* (EEFIC) bezeichneten fünf Stämme eine eigene monophyletische Gruppe in Klade F (siehe Abb.5.3). Charakteristisch für EEFIC-Stämme ist die Produktion von exfoliativen Toxinen und die häufige Assoziation mit dem *spa*-Typ t171. Darüber hinaus besitzen sie eine in das Genom integrierte Pathogenitätsinsel mit dem Namen SaRI<sub>*fusB*</sub>, die das Gen *fusB* beinhaltet und EEFIC-Stämmen eine Resistenz gegenüber Fusidinsäure verleiht [116]. Die geographische und zeitliche Ausbreitung von EEFIC-Stämmen wurde bereits für einige europäische Länder wie Holland, Großbritannien, Frankreich und Irland untersucht [127, 78]. Aus Abb.5.3 wird deut-

lich, dass sich dieser Klon seit mindestens 2005 auch in Deutschland und Tschechien verbreitet hat.

Auch muss im Zusammenhang mit phylogeographischen Analysen berücksichtigt werden, dass ein epidemiologischer Zusammenhang zwischen Isolaten durch die zum Teil geringe Probengröße pro Standort nicht immer ausgeschlossen werden kann. Trotz dieser Einschränkungen sind insbesondere für Isolate aus Ozeanien, Afrika und Mittelamerika überregionale Zusammenhänge erkennbar. Beispielsweise beinhaltet Klade A fast ausschließlich – mit Ausnahme eines Isolats aus Nordamerika (Haplotyp 10) – Isolate aus Ozeanien (Abb. 5.4 und 5.3). Der phylogenetische Zusammenhang dieser Isolate innerhalb der Klade zeigt, dass diese Isolate trotz der Asservierung in verschiedenen klinischen Labors in unterschiedlichen Regionen (Samoa, Tonga, Fidschi und Neuseeland) gemäß ihres geographischen Ursprungs Cluster bilden. Ähnliches lässt sich auch für Isolate aus Afrika und Mittelamerika ableiten.

### 6.1.3 Phylogenetische Relevanz von *spa*-Typen

Um Aufschluss über die Stabilität von *spa*-Typen bei populationsgenetischen Fragestellungen zu erlangen, wurde im Rahmen der vorliegenden Arbeit deren phylogenetische Verteilung innerhalb der CC121-Population untersucht. Die Ergebnisse zeigen eine ausgeprägte Heterogenität bei der phylogenetischen Verteilung der einzelnen *spa*-Typen (Abb. 5.9-I und -II). Insgesamt konnten 38 verschiedene *spa*-Typen identifiziert werden. Während der Großteil der identifizierten *spa*-Typen (21) nur ein Mal auftraten (*Singletons*), kamen sechs andere *spa*-Typen in 67 % aller CC121-Isolate vor. Einige der *spa*-Typen konnten mit geographischen Gebieten assoziiert werden, während andere unabhängig voneinander in mehreren phylogenetischen Sublinien auftraten. Für die drei *spa*-Typen t314, t645 und t940 konnte mit Hilfe der Software BaTS eine signifikante Korrelation mit bestimmten phylogenetischen Linien ermittelt werden (Tab. 5.6). Beispielsweise stammten alle Isolate mit dem *spa*-Typ t314 aus Klade B, die sich hauptsächlich durch afrikanische Isolate definiert. Diese Befunde decken sich mit aktuellen Studien, in denen vermehrt über afrikanische Isolate mit dem *spa*-Typ t314 als prävalenter *lukS/F-PV*-positiver MSSA-Klon berichtet wird [11]. Die globale Verbreitung von *spa*-Typ t159 lässt vermuten, dass es sich bei diesem *spa*-Typ um den Vorläufer handelt, aus dem sich nachfolgend im Laufe der Evolution unabhängig voneinander alle weiteren *spa*-Typen dieses Klon entwickelt haben.

Homoplasien konnten bei den *spa*-Typen t170, t272, t284 und t2155 detektiert werden (Abb. 5.9-II). Das Auftreten von Homoplasien belegt, dass die epidemiologische Einordnung von *S. aureus* nicht allein auf der Basis der Differenzierung von *spa*-Typen erfolgen sollte [111]. *Spa*-Sequenzen sind durch ihre Zusammensetzung aus Tandem-Repeat-Sequenzen sehr variabel und sehr anfällig für den Einbau

von Fehler während der DNA-Replikation. Schleifenbildungen in der Repeatregion können zu Deletionen oder Insertionen ganzer *Repeats* führen [95]. Diese Mechanismen sind Voraussetzung für die schnelle Evolution von *spa*-Typen und verantwortlich dafür, dass identische *spa*-Sequenzen im Verlauf der Evolution offenbar mehrfach und unabhängig voneinander entstanden sind [111]. Die ausschließliche Interpretation von *spa*-Typen bildet daher die geographische Verbreitung individueller Klone möglicherweise fehlerhaft ab. Dies konnte bereits bei Studien zu verschiedenen *S. aureus*-Linien, wie ST5, ST225 und ST22, beobachtet werden und sollte insbesondere beim Vergleich von Isolaten unterschiedlicher geographischer Herkunft berücksichtigt werden [111, 110, 159]. Die Vielzahl verschiedener *spa*-Typen, die in der CC121-Population detektiert werden konnte, lässt eher auf eine endemische Langzeit-Persistenz als auf eine klonale Ausbreitung bestimmter *spa*-Typen schließen.

Die *spa*-Typisierung gilt zwar aktuell als Basismethode zur molekularen Typisierung von *S. aureus*. Dennoch zeigen die vorliegenden Ergebnisse und die anderer Studien, dass die Zuordnung von Isolaten zu klonalen Linien (ST) und klonalen Komplexen (CC) aufgrund des *spa*-Typs bei epidemiologischen Analysen über begrenzte Zeiträume oder für geografische Regionen je nach Fragestellung und Sachlage durch weitere molekulare Typisierungsmethoden (z.B. MLST) ergänzt werden sollte [61, 111].

#### 6.1.4 Evolution klinisch unterscheidbarer Subpopulationen

*S. aureus* gehört zu den wichtigsten Infektionserregern des 21. Jahrhunderts. Vor allem CC121 ist eine der vorherrschenden klonalen Linien, die besonders bei Patienten ohne Risikofaktoren für eine **Health-Care-Associated-Infection**-(HAI) als einer der Hauptverursacher von Infektionen des Haut- und Weichteilgewebes gilt [21, 126, 151, 77]. Die immense Bedeutung dieses Klons wird in der Studie von Masiuk *et al.*, 2010 deutlich, die nachweisen konnten, dass 55 % aller Furunkulose-Fälle, die in Polen zwischen 2002 und 2008 durch MSSA-Isolate verursacht wurden, mit CC121 assoziiert waren [93].

Innerhalb der CC121-Population korrelierte nicht nur die phänotypische Verteilung des klinischen Krankheitsbildes (oberflächliche und tiefgehende Infektionen), sondern auch die drei untersuchten Virulenzdeterminanten ETA, ETB und PVL (Abb.5.6 und 5.7) mit der SNP-basierten Phylogenie (Abb.5.5-II). Der Vergleich dieser beiden Ergebnisse (Abb.5.5-II und 5.7) zeigt, dass die Ausprägung der beiden definierten Krankheitsbilder sich auf die genetische Komposition dieser drei Virulenzfaktoren bezieht. Wie Abbildung 5.5-II zeigt, sind alle tiefgehenden Infektionen mit den Virulenzgenen *lukS/F-PV* assoziiert und außerhalb von Klade F in mehreren phylogenetischen Sublinien manifestiert (siehe Abb.5.7), während alle oberflächlichen Hautinfektionen in Verbindung mit Klade F und den Exfoliatinen A und B oder,

beispielsweise bei Isolat 00-01004-1 nur mit ETA vorkommen. Die drei Isolate 00-01004-1, 97-01548 und 97-01848-1 aus Klade F wurden vor einigen Jahren im Rahmen der Routinetätigkeit des NRZs für Staphylokokken positiv für das ETB-Gen getestet (geht aus der NRZ-Datenbank hervor). Dieser Befund konnte im Rahmen der vorliegenden Studie nicht validiert werden, da hier alle drei Isolate nur bei der ETA-Analyse positiv reagierten. Die genomische Varianz liegt vermutlich in der langen Archivierung und der wiederholten Subkultivierung dieser Stämme begründet, so dass es aufgrund der plasmidalen Kodierung zum Verlust des ETB-Gens gekommen sein könnte [154]. Im Unterschied dazu liegt das Gen, welches für ETA kodiert, auf dem Bakterienchromosom und zeigt daher eine höhere genetische Stabilität [42]. Bemerkenswert an diesen Ergebnissen ist, dass bestimmte klinische Phänotypen, die mit distinkten Virulenzfaktoren assoziiert sind, sich phylogenetisch in bestimmten Sublinien manifestiert haben und signifikant mit der bestehenden Phylogenie von CC121 korrelieren. Diese Zusammenhänge wurden zwar bereits mit der Verwendung hochauflösender Typisierungsmethoden vermutet [144], konnten aber für die Spezies *S. aureus* bisher nicht bewiesen werden. Damit sind spezifische Veränderungen im akzessorischen Genom für unterschiedliche phänotypische Charakteristika – wie tiefgehende oder oberflächliche Infektionen – verantwortlich. Die Folgen der entsprechenden Infektionen können verschiedenartig ausgeprägt sein, obwohl alle Isolate den gleichen MLST-Genotyp teilen.

Wie die vorliegenden Daten demonstrieren, existieren verschiedene *S. aureus*-Stämme mit spezifischen klinischen Krankheitsbildern innerhalb eines klonalen Komplexes. Diese phylogenetische Differenzierung wurde vor allem durch die bessere Trennschärfe der genomweiten SNP-Analyse ermöglicht. Durch die bisher verwendeten molekularen Typisierungsmethoden (MLST, *spa*, PFGE) war es zwar möglich Verwandtschaften zu ermitteln und Isolate in einen klonalen Komplex (z.B. CC121) einzugruppieren, aber das geringe Auflösungsvermögen dieser Methoden erlaubte es nicht, mögliche Rückschlüsse auf Beziehungen innerhalb dieser klonalen Linien zu ziehen oder Informationen über das jeweilige Virulenzpotenzial eines Isolates abzuleiten [37]. Da die genetischen Grundlagen, die diese Unterschiede innerhalb einer klonalen Linie verursachen, nicht an die entsprechenden MLST-Daten geknüpft sind [144]. In diesem Kontext ist daher der Einsatz von Hochdurchsatz-Sequenzierungstechnologien und der sich daraus ergebenden Möglichkeit, schnell und kostengünstig vollständige Genom-Sequenzierungen durchführen zu können unverzichtbar, um ein höheres Auflösungsvermögen und eine bessere phylogenetische Differenzierung von Bakterienpopulationen zu ermöglichen [144].

### 6.1.5 Untersuchung zur Sequenzvariation innerhalb der *lukS/F-PV*-Gene

Die Analyse der *lukS/F*-Gene (1,7 kb) ergab neben den bereits beschriebenen PVL-Genotypen R1, R2, H1, H2 und H3 [13, 113], noch sechs neue Genotypen, die mit Neu-I bis Neu-VI bezeichnet wurden (siehe Tabelle 5.5).

Die phylogenetische Verteilung der unterschiedlichen PVL-Genotypen innerhalb der CC121-Population zeigt, dass Isolate mit dem PVL-Genotyp H2 über die gesamte phylogenetische Gruppe der *lukS/F-PV*-positiv getesteten Isolate verteilt sind (Abbildung 5.8). Außerdem fällt auf, dass der PVL-Genotyp R bis auf eine Ausnahme (Isolat 10-01754) ausschließlich in Klade E vorkommt und mit Isolaten asiatischer Herkunft assoziiert ist. Dies steht im Widerspruch zu den Hypothesen von O'Hara *et al.*, 2008 welche behaupten, dass einzelne PVL-Genotypen spezifisch für bestimmte geographische Regionen seien [113]. Bisher wurde angenommen, dass Isolate mit den PVL-Genotypen H1 und H2 vorwiegend in Südafrika und Indien auftreten, während Isolate mit dem PVL-Genotyp H3 hauptsächlich in Europa vorkommen [113]. Des Weiteren sollten Isolate mit dem PVL-Genotyp R überwiegend in den USA verbreitet und oft mit dem Vorhandensein von *mecA* assoziiert sein [113, 13]. Auch Letzteres konnte nicht bestätigt werden, da lediglich ein MRSA-Isolat in Klade F (11-02369) neben sieben weiteren MSSA-Isolaten mit dem PVL-Genotyp R2 identifiziert werden konnte (Abb.5.8). Lediglich PVL-Genotyp H3 ist vorwiegend mit europäischen Isolaten assoziiert (Ausnahme bilden die zwei Isolate 10-01142 und 10-01735) und hauptsächlich in Klade D organisiert, was den Angaben von O'Hara *et al.* 2008 entspricht (Abb.5.8).

Die deutlich höhere Diversität und die globale Verbreitung von Isolaten des PVL-Genotyps H2 in der untersuchten CC121-Population decken sich mit den Annahmen vorheriger Studien. In denen wird davon ausgegangen, dass der PVL-Genotyp H2 ein Ausgangstyp darstellt, der sich global ausgebreitet hat und aus dem im Laufe der Evolutionsgeschichte weitere *lukS/F-PV*-Varianten hervorgegangen sind [13, 113]. Unterstützt wird diese Aussage durch die weite Verbreitung von H2 in verschiedenen klonalen Komplexen und der relativ großen Sequenzvariabilität gegenüber dem PVL-Genotyp R mit seinen Varianten R1 und R2 (siehe Tabelle 5.5) [13, 142].

Generell sollte der Versuch der Verknüpfung bestimmter PVL-Genotypen mit bestimmten epidemiologischen Fragestellungen und einer geographischen Manifestierung nur vorsichtig erfolgen, da die von O'Hara vorgeschlagene Einteilung in PVL-Genotypen nur auf wenige Sequenzunterschiede in den Toxingenen *lukS/F-PV* beruht und die wirkliche Diversität der vermittelnden Prophagen-Population nicht berücksichtigt wird. Durch die Genom-Sequenzierung von sieben PVL-positiv getesteten CC121 Isolaten, die im Rahmen der Genomanalysen ausgewählt wurden, und den anschließenden Genomvergleichen, war es möglich einen Einblick in die Diversität der PVL-tragenden Phagengenome zu erhalten (siehe Abb.5.14). Besonders interes-



sant war der Vergleich der kompletten Prophagen-Genomsequenzen der vier Isolate 10-00259, 11-00380, 11-02371 und 11-02376, die aufgrund von Sequenzpolymorphismen im PVL-Lokus dem PVL-Genotyp H2 zugeordnet können (Tab.5.5). Die Sequenzen zeigen zum Teil große strukturelle Ähnlichkeiten mit einzelnen SNPs (von 6 bis >750 SNPs), es kommen aber auch Regionen vor, die erhebliche Sequenzunterschiede aufweisen. Auffällig waren die Sequenzhomologien der beiden Prophagen in den asiatischen MRSA-Isolaten 11-02371 und 11-02376, die sich nur um sechs SNPs unterscheiden, was die enge phylogenetische Verwandtschaft unterstreicht (Abb.5.8). Ebenfalls kritisch zu sehen ist der Punkt, dass die Einteilung der verschiedenen PVL-Genotypen nicht ausschließlich auf nicht-synonyme Mutationen, die einem gewissen Selektionsdruck unterliegen, sondern auch auf synonymen Mutationen (Position 282, 367, 663, 1186 und 1396, siehe Tab. 5.5) beruht. Diese unterliegen keinem Selektionsdruck und reichern sich an Zufallspositionen an. Die Nukleotidvariationen im *lukS/F-PV*-Gen und die hohe Diversität unterschiedlicher PVL-Prophagen innerhalb der CC121-Population geben Aufschluss darüber, dass es im Laufe der evolutionären Entwicklung von CC121 mehrfach zur unabhängigen Akquirierung von PVL-tragenden Prophagen gekommen ist. Anhand der phylogenetischen Analysen und der Genomvergleiche der Prophagen können mindestens 14 unabhängige Aufnahmeereignisse von PVL tragenden Prophagen identifiziert werden (siehe Abb. 5.8).

### 6.1.6 Ursachen für die evolutionäre Divergenz zweier Teilpopulationen

Wie die Ergebnisse der vorliegenden Arbeit zeigen, gibt es innerhalb der CC121-Population zwei Subpopulationen mit unterschiedlichen klinischen Phänotypen. Diese Tatsache wirft die Frage auf, warum es innerhalb dieser Population zum gegenseitigen Ausschluss von *lukS/F-PV* und *eta*-vermittelnden Prophagen kam. Es war in dieser Studie nicht möglich, Isolate zu detektieren, die beide Exotoxine (PVL und Exfoliatine) bilden. Bis dato liegen auch keine Ergebnisse in der Literatur vor, bei denen eine solche Koexistenz in *S. aureus* beschrieben wurde. In einer Studie von Goerke *et al.*, 2009 wurde ein Klassifizierungsschema für die in der *S. aureus*-Population vorkommenden Prophagen etabliert, welches auf der Nukleotidsequenz des Integrase-Gens beruht [47]. 24 der 34 identifizierten Prophagen konnten anhand ihrer Integrase-Gene in vier Gruppen unterschieden werden (Tab.5.9). All diese Prophagen integrieren an unterschiedlichen Positionen (*attachment sites*).

Damit kann ausgeschlossen werden, dass ein im Genom integrierter Prophage von einer Integrase-Gruppe die Aufnahme eines zweiten Phagen, einer anderen Integrase-Gruppe, durch die Blockierung der spezifischen *attachment site* verhindert. Wie es zu einer Beeinträchtigung des *lukS/F-PV* vermittelnden Prophagen durch ETB kommt, bleibt ungewiss, da ETB nicht chromosomal vererbt wird, sondern plasmidal

kodiert ist und bisher keine Informationen über eine Wechselwirkung mit Prophagen zur Verfügung stehen.

Der PVL-Prophage der Integrase-Gruppe Sa2 integriert in seine *attachment site* im ORF SA1320 (ORF-Bezeichnung in Isolat N315) [65]. Dieses Gen einschließlich der *attachment site* ist im Isolat 09-00227 deletiert. Da das Vorhandensein der 29 bp langen *attachment site* Voraussetzung für eine erfolgreiche Integration ist [65], lag die Vermutung nahe, dass die unterschiedliche evolutionäre Entwicklung von zwei klinischen Pathotypen aus dem Verlust der spezifischen *attachment site* resultiert. Die Voraussetzungen für eine erfolgreiche Übertragung eines PVL-Prophagen wären damit nicht mehr gegeben. Unterstützt wird diese Beobachtung durch eine Arbeit von Wirtz *et al.* 2010, in der ein ungewöhnliches Ausschneiden eines PVL-kodierenden Prophagen aus einem ST80-Isolat beobachtet werden konnte [158]. Hierbei kommt es zu einer intrachromosomalen Rekombination innerhalb eines *Tandem Repeat Cluster* [143], welches den ORF SA1320 mit dem integrierten Prophagen umgibt. Bei diesem *Crossing-over* zwischen den paralogenen Genen wird chromosomale DNA des Bakteriums mit in den Prophagen übertragen [158]. In der vorliegenden Untersuchung war vermutlich ebenfalls ein *Crossing-over* zwischen ORF SA1319 und SA1321 in Isolat 09-00227 die Ursache für diese intrachromosomale Rekombination und somit der Grund für die Deletion von ORF SA1320 samt *attachment-site*. Um dies auch für die anderen Isolate in Klade F zu bestätigen, wurde eine Nachweis-PCR für die spezifische *attachment site* des PVL-Prophagen entwickelt. Fast alle Isolate in Klade F – bis auf die Isolate der monophyletischen Gruppe der EEFIC-Stämme (siehe Abb.5.3) und das Isolat 10-01149 – zeigten die für eine erfolgreiche Integration des PVL-Prophagen benötigte *attachment-site*. Somit kommt das Fehlen der *attachment-site* nicht als Ursache für die Evolution der klinisch unterscheidbaren Subpopulationen in Frage.

Möglicherweise ist die beobachtete Assoziation der untersuchten Virulenzdeterminanten mit dem klinischen Phänotyp und die Korrelation mit bestimmten phylogenetischen Sublinien eher durch die flächendeckende Ausbreitung einiger weniger klonaler Stämme mit einer der betreffenden Virulenzdeterminanten zu erklären (siehe Abbildung 5.3 und 5.7), als durch spezifische genomische Mechanismen, die eine zusätzliche Akquirierung von Sa1- oder Sa2-Prophagen unterbinden. Liegen solche Mechanismen vor, sollte es Hinweise dafür geben, dass auch Prophagen anderer Integrase-Gruppen davon betroffen sind. Dagegen spricht wiederum die hohe Diversität von 34 Prophagen und deren Einteilung in vier Integrase-Gruppen, welche innerhalb der untersuchten Stammauswahl identifiziert werden konnten (Tab.5.9). So war es möglich in allen sequenzierten CC121-Isolaten im Mittel zwei bis vier Prophagen unterschiedlicher Integrase-Gruppen zu identifizieren, die stabil ins Genom integriert waren.

Eine andere denkbare Ursache für die gesteigerte epidemiologische Fitness der SNP-basierten Kladen und deren sich klonal ausbreitenden Stämme könnten auch ihrer

Antibiotika-Resistenzprofile sein. Mit Hilfe der Software BaTS konnte ausschließlich in den Isolaten in Klade B für die Resistenz gegenüber dem Antibiotikum Sulfamethoxazol/Trimethoprim (SXT) eine signifikante Korrelation ( $p = 0.01$ ) zur Phylogenie ermittelt werden (siehe Tab.A.12). Dieses Ergebnis deckt sich mit bereits publizierten Studien [44, 115]. Sehr häufig wird bei Stämmen, sowohl von besiedelten Probanden als auch aus Infektionsgeschehen, mit afrikanischem Ursprung eine Antibiotikaresistenz gegenüber SXT beobachtet [44, 115]. Dieser Befund kann nicht allein durch den übermäßigen und unkontrollierten Einsatz dieses Antibiotikums in Afrika erklärt werden und ist Gegenstand aktueller Forschungen [44]. Alle anderen nachgewiesenen Antibiotikaresistenzen zeigten keine signifikante Korrelation mit bestimmten phylogenetischen Gruppen (siehe Tab.A9-11,13-16), so dass eine Ursache für die Assoziation klinischer Phänotypen und Virulenzdeterminanten mit bestimmten SNP-basierten Kladen nicht durch das Antibiotika-Resistenzprofil erklärt werden kann.

## 6.2 Vergleichende Genomanalysen

Um einen Überblick über die genomische Variabilität in der untersuchten CC121-Population zu erhalten, wurden zehn repräsentative Isolate ausgewählt und unter Verwendung der *Next-Generation-Sequencing*-(NGS)-Technologie 454 vollständig sequenziert. Wie bereits für *S. aureus* beschrieben [84], weist auch das Kerngenom in der untersuchten CC121-Population eine geringe Variabilität auf. Die Diversität der einzelnen Isolate beruht überwiegend auf verschiedene mobile genetische Elemente, die das akzessorische Genom in den untersuchten *S. aureus*-Isolaten bilden. Insgesamt konnten in den zehn sequenzierten Genomen 31 Pathogenitätsinseln und 34 Prophagen identifiziert werden (Tab.5.9).

### 6.2.1 *S. aureus* Pathogenitätsinseln (SaPIs)

Alle identifizierten Pathogenitätsinseln lassen sich aufgrund ihrer Integrationsstelle im Genom und auf Basis der Integrase-Homologie in vier von sechs bisher beschriebenen SaPI-Gruppen einordnen [108]. Pathogenitätsinseln sind chromosomale Strukturen, die vermehrt Virulenzgene tragen und meist in Bakteriengenomen lokalisiert sind. Ihre Bedeutung für die Evolution von Gram-positiven Pathogenen wurde bereits erkannt [9, 30]. 2007 waren für *S. aureus* 20 verschiedene SaPIs sequenziert worden, die aufgrund unterschiedlicher *attachment sites* (*attC*) in sechs Gruppen unterteilt werden [108].

Innerhalb der einzelnen SaPI-Gruppen zeigen die identifizierten Pathogenitätsinseln bis auf SaPI1 nur eine geringe Diversität (Tab.5.9). Die wenig variablen SaPI2 und SaPI3 kommen in allen der zehn untersuchten Isolate vor. Die Pathogenitätsinsel SaPI4 konnte in sechs Isolaten identifiziert werden und zeigte ebenfalls eine nur

geringe Diversität (Tab.5.9). In fünf Isolaten konnte eine SaPI1 identifiziert werden (Tab.5.9), die sich bezüglich des Gengehaltes unterschieden. So konnte in dem Isolat 10-00380 das Gen für das Superantigen Toxisches-Schock-Syndrom-Toxin 1 (TSST-1) auf der ca. 27 kb großen SaPI1 detektiert werden. Während die Pathogenitätsinseln in den drei Isolaten 98-01618, 09-02483 und 09-02298 relativ konserviert sind und keine bekannten Virulenzdeterminanten tragen, konnte für das Isolat 09-00227 die für EEFIC-Stämme charakteristische *S. aureus*-Resistenzinsel *SaRI<sub>fusB</sub>*, die zur Fusidinsäure-Resistenz beiträgt, detektiert werden [116].

Auffällig war, dass in allen der zehn sequenzierten Genomen nur jeweils eine Kopie des *hsdS*-Gens in SaPI2 identifiziert werden konnte (Tab.5.9). Dieser Befund unterscheidet die sequenzierten Isolate von anderen klonalen Linien, bei denen das Gen immer Teil der Pathogenitätsinsel SaPI2 und SaPI3 ist [6]. Das Gen *hsdS* kodiert für eine der drei Untereinheiten (die DNA erkennende Untereinheit) des RM-Systems das Waldron *et al.* als Sau1 Typ-I RM-System in *S. aureus* beschrieben haben [6, 152]. Dieses System steht im direkten Zusammenhang mit der Abwehr von Fremd-DNA und der Blockierung des horizontalen Gentransfers zwischen verschiedenen *S. aureus*-Stämmen [152]. Das Fehlen des *hsdS*-Gens in SaPI3 bei den untersuchten Isolaten könnte den horizontalen Genaustausch von MGEs beeinträchtigen und damit die genetische Variabilität dieser Population mitbestimmen. Dennoch bleibt die tatsächliche Ursache unbekannt und die Bedeutung des RM-Systems sollten vorsichtig interpretiert werden.

Generell konnte die größte Diversität bei den Pathogenitätsinseln in der SaPI1-Gruppe beobachtet werden, wohingegen die anderen Pathogenitätsinseln der SaPI-Gruppen 2-4 wenig variabel waren. Die Annahme liegt nahe, dass SaPI2 und SaPI3 schon früh in der Evolutionsgeschichte der CC121-Population aufgenommen wurden und mit ihr koevolvierten. Im Gegensatz dazu stehen SaPI4 und SaPI1, die offenbar im Verlauf der Evolution durch Helferphagen mehrfach unabhängig voneinander übertragen wurden [134].

### 6.2.2 Prophagen

Bei der Analyse der in das Genom integrierten Prophagen wurde mit der Identifizierung von 34 Prophagen die stärkste Diversität beobachtet (Tab.5.9). Der Großteil der Prophagen konnte aufgrund des von Goerke *et al.* 2009 etablierten Klassifizierungsschemas in die vier verschiedenen Integrase-Gruppen Sa1, Sa2, Sa3 und Sa5 eingeteilt werden [47]. Die Ergebnisse legen durch die starke Diversität der Prophagen eine lange evolutionäre Entwicklungsgeschichte von CC121 nahe. Auch die ausgedehnten Sequenzpolymorphismen, die im PVL-Lokus identifiziert werden konnten und die resultierende Einteilung in die verschiedenen PVL-Genotypen, unterstützt diese Annahme. Neben den über die Integrase-Gruppe klassifizierten Prophagen konnte ein weiterer, bisher unbekannter, Prophage ( $\Phi$ SaCC121) identifiziert wer-

den. Dieser liegt mit einer Größe von ca. 41 kb in allen zehn untersuchten Isolaten an immer der gleichen Stelle und stark konserviert im *S. aureus*-Genom integriert vor (N315 ORF-Bezeichnung; zwischen SA1150 und SA1154). Über den möglichen Einfluss des Prophagen auf die CC121-Population kann im Rahmen dieser Arbeit nur spekuliert werden. Die BLAST-Ergebnisse zeigten neben typischen Phagengenen (funktionelle Gene der Morphogenese, DNA-Verpackung, Integration oder Exzision des Prophagen) keine Virulenz- oder Resistenz-assoziierten Gene, die der CC121-Population einen Vorteil gegenüber anderen klonalen Linien schaffen würden.

Interessant ist die Region, in der  $\Phi$ SaCC121 integriert, da der Bereich zwischen den ORF SA1150 und SA1154 im Vergleich zu anderen *S. aureus*-Genomen unterschiedlicher klonaler Linien stark in seiner genomischen Zusammensetzung variiert. Diese Region ist dem Anschein nach einer ständigen genetischen Umstrukturierung unterworfen, da vor allem trunkierte Strukturen mobiler genetischer Elemente gefunden werden können, die vermutlich für das Auftreten von Translokationen, Inversionen und anderen chromosomalen Veränderungen verantwortlich sind. Umso bemerkenswerter ist, dass  $\Phi$ SaCC121 in allen untersuchten Isolaten, ob aus Europa, Afrika, Asien, Südamerika oder Ozeanien auftrat. Dies lässt den Schluss zu, dass sich  $\Phi$ SaCC121 aus evolutionsbiologischer Sicht schon sehr früh in der CC121-Population etablierte, damit ein Relikt aus der ursprünglichen Population darstellt und vermutlich sogar mit ihr koevolvierte. Welchen evolutionären Vorteil  $\Phi$ SaCC121 für die CC121-Population bietet, konnte in dieser Arbeit nicht geklärt werden. Auch die Frage warum  $\Phi$ SaCC121 im Zuge der natürlichen Selektion nicht aus der Population entfernt wurde, bedarf weiterer Analysen.

## 6.3 Die Bedeutung von CC121-MRSA

Auffällig an CC121 ist vor allem die Tatsache, dass es in dieser *S. aureus*-Linie bis 2009 nie zu einer Integration einer *SCCmec*-Kassette gekommen war [28]. Erst seit 2009 erschienen einzelne Publikationen ausschließlich aus dem asiatischen Raum (insbesondere China, Kambodscha und Japan), in denen erstmals ST121-Isolate als MRSA identifiziert wurden [16, 103, 38, 70]. Durch die vergleichenden Genomanalysen und die phylogenetischen Analysen, die auf 1,7 % des *S. aureus*-Genoms basierten, war es im Rahmen der vorliegenden Arbeit möglich, einen Einblick in die Abstammungsverhältnisse der sich dort entwickelnden MRSA-Population zu bekommen und Vergleiche mit vorhandenen MSSA-Isolaten auf der Genomebene durchzuführen.

### 6.3.1 Phylogenetische Beziehung zwischen den MRSA- und MSSA-Isolaten

Sowohl die phylogenetischen Analysen als auch die Analysen, die auf den variablen Positionen des Kerngenoms der zehn sequenzierten Isolate basieren, zeigen, dass die untersuchten asiatischen MRSA-Isolate in eine monophyletische Gruppe (Klade E) geordnet werden können (siehe Abb.5.3 und 5.12). Auch alle MSSA-Isolate aus dieser Region clustern in Klade E und sind laut der Rekonstruktionen phylogenetisch älter als die MRSA-Isolate (siehe Abb.5.3 und 5.12). Vor allem die rekonstruierten phylogenetischen Stammbäume in Abbildung 5.12, die auf 1828 SNPs des Kerngenoms in den zehn sequenzierten Isolaten basieren, zeigen eine sehr enge phylogenetische Verwandtschaft zwischen den beiden MRSA-Isolaten 11-02369 und 11-02371, die sich von dem basal liegenden MSSA-Isolat 11-02376 phylogenetisch abtrennen. Die bedingte Autapomorphie der SCC*mec*-Kassette innerhalb dieser Klade macht die Entstehung der auftretenden MRSA-Isolate aus der vorherrschenden MSSA-Population wahrscheinlich. Das annähernd identische Antibiotogramm, das zwischen den MSSA- und den MRSA-Isolaten in dieser Region vorherrscht, unterstützt diese Annahme [16].

### 6.3.2 Genomvergleiche

Lange wurde vermutet, dass CC121-Isolate genomischen Mechanismen unterliegen, die eine Akquirierung von SCC*mec*-Elementen verhindern bzw. deren stabile Integration in das Genom unterdrücken [28]. Deshalb sollten mit Hilfe von Ganzgenom-Analysen diese postulierten genomischen Mechanismen aufgeklärt werden. In diesem Kontext wurden mögliche funktionelle Unterschiede zwischen CC121-MRSA- und CC121-MSSA-Isolaten in der *orfX*-Region mit der spezifischen *attB*<sub>SCC</sub>-Sequenz oder in der Umgebungssequenz der *orfX*-Region vermutet.

Die Analysen der *orfX*-Region mit der spezifischen *attB*<sub>SCC</sub>-Sequenz, die zur Akquirierung des SCC*mec*-Elementes erforderlich ist [62], zeigten keine Abweichungen zwischen den untersuchten MRSA- und MSSA-Isolaten. Somit ist das Fehlen der sequenzspezifischen Integration eines SCC*mec*-Elementes durch die Rekombinase-Systeme CcrAB und CcrC nicht auf eine Irregularität in der *orfX*-Region zurückzuführen.

Die Untersuchung der DNA-Umgebung der spezifischen *attB*<sub>SCC</sub>-Sequenz basiert auf einer Studie von Noto *et al.* (2007) [107], die sich vorwiegend mit der Integration und Excision von SCC*mec* in bestimmten *S. aureus*-Linien beschäftigten. Ihre besondere Aufmerksamkeit galt dabei dem Isolat MW2 (Akzessionsnummer: BA000033). MW2 wies trotz integrierter SCC*mec*-Kassette eine ähnliche Umgebungssequenz

der *orfX*-Region auf, wie sie sonst nur bei typischen MSSA-Linien (darunter auch ein ST121-Isolat 3298) beobachtet werden konnte [107]. Die experimentellen Bemühungen, eine Rekombinase-vermittelte Exzision der SCC*mec*-Kassette einzuleiten, scheiterten im Gegensatz zu Experimenten an anderen klonalen MRSA-Linien [106]. Noto *et al.* schlussfolgerten daraus, dass die Umgebungssequenz der *orfX*-Region einen entschiedenen Einfluss auf die sequenzspezifische Integration eines SCC*mec*-Elementes bzw. den Rekombinase-vermittelten Mechanismus hat [107]. Die Genomvergleiche der zehn in der vorliegenden Studie sequenzierten Isolate zeigten aber auch in der spezifischen Umgebungssequenz keine Irregularität zwischen den untersuchten MRSA- und MSSA-Isolaten. Dennoch war ein interessanter Aspekt, dass alle sequenzierten Isolate die gleiche charakteristische Umgebungssequenz aufwiesen, die zuvor nur bei typischen MSSA-Linien beobachtet werden konnte [107]. Die von Noto *et al.* (2007) publizierte Umgebungssequenz (aus dem ST121-Isolat 3298) wies neben einer Punktmutation in den Isolaten 11-02369, 11-02371 und 11-02376 keine weiteren Unterschiede zwischen den untersuchten MRSA- und MSSA-Isolaten auf [107]. Ob diese Punktmutation oder die charakteristische Umgebungssequenz der *orfX*-Region einen Einfluss auf die SCC*mec*-Aufnahme hat, konnte im Rahmen dieser Arbeit nicht geklärt werden.

Für die Beantwortung der Frage, warum es in CC121 offenbar nur selten (erst 2009) zu einer Akquirierung von SCC*mec* kommt, müssen auch Mechanismen unabhängig von den bekannten Rekombinase-Systemen berücksichtigt werden. Noto *et al.* (2006) vermuteten aufgrund der fehlgeschlagenen Bemühungen, in Isolat MW2 eine Rekombinase-vermittelte Exzision der SCC*mec*-Kassette einzuleiten, dass die Aufnahme des SCC*mec*-Elementes durch eine homologe Rekombination und nicht durch die üblichen Rekombinasen, kodiert durch die Rekombinase-Gene *ccrA*, *ccrB* und *ccrC*, erfolgte [106]. Für die MRSA-CC121 Population ist keine fundierte Erklärung möglich, da weitere Analysen hierfür erforderlich sind.

In diesem Kontext soll auch das RM-System diskutiert werden. Das RM-System stellt eines der bedeutendsten Systeme zur Abwehr von Fremd-DNA dar [99]. Daher ist es möglicherweise auch in der Lage die Akquirierung von SCC*mec* in zu verhindern. Auffällig bei den Untersuchungen zur Struktur der Pathogenitätsinseln war, dass die Gen-Determinante *hsdS* des weitverbreiteten SauI Typ-I RM-Systems in allen sequenzierten Isolaten innerhalb der Pathogenitätsinsel SaPI3 trunkiert vorlag. Dies stellt eine andere Situation dar als in anderen klonalen *S. aureus*-Linien, bei denen *hsdS* immer ein Teil der Pathogenitätsinseln SaPI1 und SaPI2 ist [6]. Waldron und Lindsay äußerten die Vermutung, dass sich die Variation des *hsdS* Gens gravierend auf den horizontalen Gentransfer innerhalb einer klonalen Linie auswirkt und möglicherweise eine Rolle bei der Entwicklung einer klonalen Populationsstruktur von *S. aureus* spielt [152]. Die Deletion des *hsdS* Gens – und der damit verbundenen Reduzierung von zwei Kopien auf eine – könnte die Variabilität des RM-Systems

bei der Auswahl spezifischer DNA-Sequenzen innerhalb der CC121-Population und damit den Austausch von mobilen genetischen Elementen limitieren. Gegen diese Theorie spricht aber die Tatsache, dass bei den Genomanalysen in der vorliegenden Studie mit einer Gesamtzahl von 34 Prophagen und 31 Pathogenitätsinseln eine relativ große Variabilität an mobilen genetischen Elementen identifiziert werden konnte. Dabei konnte im Mittel für jedes der zehn untersuchten CC121-Isolate zwei bis vier Prophagen und drei bis vier Pathogenitätsinseln detektiert werden (Tab.5.9). Diese hohe Variabilität an mobilen genetischen Elementen spricht gegen die Deletion der *hsdS* Komponente als Ursache für das seltene Auftreten von CC121 MRSA-Isolaten. Zusammenfassend konnten durch die Genomanalysen von CC121 MRSA- und MSSA-Isolaten keine Eigenschaften gefunden werden, die eine Akquirierung einer SCC<sub>mec</sub>-Kassette generell ausschließen würden. Auch das RM-System sollte als mögliche Ursache kritisch diskutiert werden, da es in *S. aureus* nicht den einzigen Mechanismus zur Kontrolle des horizontalen Gentransfers darstellt, sondern auch weitere, noch unbekannte Mechanismen dafür verantwortlich sind [150].

### 6.3.3 CC121-MRSA

CC121-MRSA sind bis jetzt nur endemisch in Asien aufgetreten (Abb.5.3) [16, 103, 38, 70]. Diese Stämme könnten aber die Entstehung eines neuen erfolgreichen MRSA-Klons darstellen. Dieses Phänomen einer erfolgreichen ökologischen Ausbreitung von bestimmten Klonen, wurde bereits für *S. aureus* beschrieben [140, 26] und auch mehrfach für MRSA-Klone, wie ST22, ST239, ST225 und ST5, dokumentiert [55, 110]. Zusätzliches Gefahrenpotential besteht durch den Erwerb des Virulenzfaktors PVL, der zu einem überproportional hohen Anteil in den MSSA-Isolaten der CC121-Population detektiert werden kann und durch die bereits sehr erfolgreiche globale Verbreitung dieses Klons als MSSA. Das große Reservoir an PVL-positiven MSSA-Isolaten könnte die Ausbreitung eines neuen, PVL-positiven CC121-MRSA-Klons begünstigen. Dies konnte bereits bei anderen Klonen beobachtet werden, bei denen ebenfalls ein Pool aus PVL-positiven MSSA-Isolaten als Reservoir für die Ausbildung PVL-positiver MRSA-Isolate diente [12, 44]. Daher sollte in Zukunft die Verbreitung derartiger Klone als wesentliche Voraussetzung für gezielte Präventionsmaßnahmen sorgfältig überwacht werden.



## 6.4 Ausblick

Im Rahmen der populationsgenetischen Analysen der vorliegenden Arbeit war es durch die erheblich verbesserte Auflösung der angewandten Technologien (dHPLC, 454-Sequenzierung) möglich, die Evolution von zwei klinisch unterscheidbaren Subpopulationen, innerhalb der klonalen *S. aureus* Linie CC121, aufzudecken. Dabei zeigte sich, dass Isolate mit bestimmten klinischen Phänotypen, die mit distinkten Virulenzfaktoren assoziiert sind, sich in spezifischen phylogenetischen Sublinien manifestiert haben und signifikant mit der bestehenden Phylogenie von CC121 korrelieren. Bisher eingesetzte Typisierungsmethoden, wie MLST oder die *spa*-Typisierung waren dazu nicht in der Lage. Im Rahmen der Arbeit konnte nicht eindeutig geklärt werden, warum es zu dieser unterschiedlichen Evolution zweier verschiedener Subpopulationen mit unterschiedlichem Virulenzpotential kam. Genomische Mechanismen, die zum gegenseitigem Ausschluss der beiden Virulenz-übertragenden Prophagen (Sa1- und Sa2-Prophage) führen, konnten als mögliche Ursachen mit hoher Wahrscheinlichkeit ausgeschlossen werden. Dennoch sollte diese Möglichkeit bei zukünftigen Projekten weiterhin berücksichtigt werden. Inwieweit die gegenwärtige Subpopulation in Klade F die gesamte CC121-Population widerspiegelt, konnte nicht eindeutig geklärt werden, da bislang mit Ausnahme von einem afrikanischen und zwei amerikanischen Isolaten, nur europäische Isolate in den Datensatz aufgenommen wurden. Weitere Isolate aus verschiedenen Kontinenten müssten in den Datensatz integriert werden, um detaillierte Rückschlüsse zur Ausbreitung und Mikroevolution dieser Subpopulation ziehen zu können.

Wie die vorliegenden Ergebnisse zeigen, ist die Analyse von genomweiten SNPs wichtig, um verwandtschaftliche Beziehungen von *S. aureus*-Stämmen zu klären und darüber hinaus die Evolution wie auch die demographischen Entwicklungen einer Population mit hoher Auflösung zu rekonstruieren. Dies stellt einen wichtigen Aspekt bei der Überwachung der Ausbreitung von *S. aureus* Populationen dar. Daher sollten für evolutionäre Fragestellungen, die eine höhere Auflösung erfordern, zukünftig Methoden mit einem hohen Diskriminierungspotential (z.B. Ganzgenom-Analysen) verwendet werden. Diese Technologien haben entscheidend zur Entschlüsselung der Populationsstruktur und Evolutionsgeschichte von CC121 beigetragen und können in Zukunft zur Einschätzung eines möglichen pathogenen Potentials eines isolierten Stammes genutzt werden. Die identifizierten SNPs (Tab.A.2) können als diagnostische Marker für bestimmte Klade assoziierte Isolate dienen und helfen phylogenetische Zusammenhänge zu entschlüsseln [76]. Des Weiteren ist die Entschlüsselung von vollständigen Genomsequenzen bei der Identifikation von Infektionsquellen und bei der Aufklärung von Transmissionswegen erfolgversprechend [55] und konnte bereits in einigen Studien eindrucksvoll bewiesen werden [137, 54]. Die vollständigen Genomsequenzen der CC121-MRSA-Isolate lieferten keine Hinweise darauf, warum es erst so spät zu einer MRSA-Entwicklung (2009) innerhalb

der CC121-Population gekommen ist. Das für die Integration der *SCCmec*-Kassette erforderliche *orfX* Gen war in allen untersuchten CC121-MSSA-Isolaten vorhanden. Auch die genomische Umgebungssequenz der spezifischen *orfX* Region zeigte keine auffälligen Unterschiede zwischen den MRSA- und MSSA-Isolaten der CC121-Population. Bei zukünftigen Analysen sollte der Blick nicht nur auf die spezifische *orfX* Region gerichtet sein, sondern weitere denkbare Mechanismen müssen mit berücksichtigt werden. Mögliche Kandidaten sind Mechanismen, die ähnlich wie das RM-System fungieren und die Aufnahme von MGEs verhindern. Des Weiteren ist vorstellbar, dass die Akquirierung von *SCCmec* unabhängig von der Rekombinase-vermittelten Integration erfolgt [106]. Mutagenese-Experimente und geeignete *in vivo* Modelle könnten bei der Aufklärung helfen, ob und in welchem Umfang genetische Unterschiede die Aufnahme von *SCCmec* beeinflussen.

Mit der vollständigen Analyse der zehn Genomsequenzen war es möglich, einen bisher unbekannten Prophagen ( $\Phi$ SaCC121) zu identifizieren.  $\Phi$ SaCC121 lag in allen untersuchten Isolaten konserviert vor. Die Frage, warum  $\Phi$ SaCC121 im Zuge der Evolution nicht aus der Population entfernt wurde, sondern mit ihr koevolvierte, konnte im Rahmen dieser Arbeit nicht geklärt werden. Es bleibt zukünftigen Analysen vorbehalten, die evolutionäre Entstehungsgeschichte dieses Prophagen und seinen Beitrag zur Fitness und globalen Ausbreitung der CC121-Population zu analysieren.

## A. Anhang

Tabelle A.1: CC121-Isolate

Isolat	NRZ-ID	Jahr	Land	spo- type	Haplotyp	MLST	SCCmec	etb	eta	luks/F	PVL	Klade	Resistenz <sup>1</sup>	Infektion
00-01004-1	148	2000	Deutschland	t159	50	ST121	—	—	x	—		F	PEN,ERY,TET	Konjunktivitis
00-01437	005 308/00	2000	Deutschland	t159	49	ST121	—	—	x	—		F	PEN,ERY	Wundinfektion
00-01876	MV 11997	2000	Deutschland	t284	51	ST121	—	—	x	—		F	PEN,ERY,TET	
02-02750	VA 041303	2002	Deutschland	t284	80	ST121	—	—	—	x	H1	D	PEN	Furunkel
02-02878	-	2002	Deutschland	t285	71	ST121	—	—	—	x	H3	D	PEN,ERY,TET	Abszess
05-02318	26240	2005	Deutschland	t435	73	ST121	—	—	—	x	H3	D	PEN	
06-01454	BK 1118	2006	Deutschland	t284	73	ST121	—	—	—	x	H3	D	PEN	Nekrotisierende Pneumonie
06-02718	240 888	2006	Deutschland	t159	57	ST121	—	x	x	—		F	PEN,FUS	Exfoliative
06-02926	123764	2006	Österreich	t159	65	ST121	—	x	x	—		F	sensibel	Dermatitis
07-00146	100452	2007	Österreich	t159	54	ST121	—	x	x	—		F	PEN,ERY	SSSS
07-00384	10000/2408480	2007	Deutschland	t171	45	ST123	—	x	x	—		F	PEN,FUS	SSSS
07-00385	VF 1524	2007	Deutschland	t314	104	ST121	—	—	—	x	H2	B	sensibel	Abszess
07-00778	C 1497 K1	2007	Deutschland	t159	60	ST121	—	x	x	—		F	PEN	SSSS
07-00888	105830	2007	Österreich	t159	55	ST121	—	x	x	—		F	PEN,ERY	SSSS
07-01428	235149	2007	Deutschland	t2375	74	ST121	—	—	—	x	H3	D	PEN	
07-01668	S522	2007	Kenia	t314	107	ST121	—	—	—	x	H2	B	PEN,TET,SXT	
07-01670	TP1	2007	Fidschi	t159	3	ST121	—	—	—	x	H2	A	PEN,ERY,CLI	SSSS
07-02764	13353	2007	Deutschland	t659	43	ST123	—	x	x	—		F	PEN,ERY,FUS	Abszess
07-03044	8519199	2007	Deutschland	t1114	73	ST121	—	—	—	x	H3	D	ERY,TET	Wundinfektion
07-03154	7K01 7003	2007	Deutschland	t1211	86	ST121	—	—	—	x	H3	D	sensibel	SSSS
08-00471	A 09372	2008	Deutschland	t160	68	ST121	—	x	x	—		F	PEN,TGC	Wundinfektion
08-00652	232303	2008	Deutschland	t435	75	ST121	—	—	—	x	H3	D	PEN	Wundinfektion
08-01906	116612	2008	Österreich	t272	61	ST121	—	x	x	—		F	PEN	Wundinfektion
08-02087	ER 12507/08	2008	Sambia	t940	37	ST121	—	—	—	x	H2	C	PEN	Abszess
08-02146	119441	2008	Österreich	t159	54	ST121	—	x	x	—		F	ERY	SSSS
09-00227	VF 21876	2009	Deutschland	t171	44	ST123	—	x	x	—		F	PEN,ERY,FUS	SSSS
09-00296	VA 40	2009	Deutschland	t159	58	ST121	—	x	x	—		F	PEN,TET	SSSS
09-00462	-	2009	Deutschland	t159	56	ST121	—	x	x	—		F	PEN	
09-00577	PA 3	2009	Panama	t940	34	ST121	—	—	—	x	H2	C	PEN,FUS	Abszess
09-00584	PA 10	2009	Panama	t940	35	ST121	—	—	—	x	H2	C	PEN,FUS	SSSS
09-01732	-	2009	Nigerien	t159	98	ST121	—	—	—	x	H2		PEN,SXT	SSSS
09-01745	-	2009	Nigerien	t2304	119	ST121	—	—	—	x	H2	B	PEN,SXT	
09-01759	-	2009	Nigerien	t314	118	ST121	—	—	—	x	H2	B	PEN,SXT	
09-01797	-	2009	Nigerien	t314	103	ST121	—	—	—	x	H2	B	TET	

Tabelle A.1: CC121-Isolate

Isolat	NRZ-ID	Jahr	Land	spo- type	Haplotyp	MLST	SCCmec	etb	eta	lukS/F	PVL	Genotyp	Klade	Resistenz <sup>1</sup>	Infektion
09-01826	-	2009	Nigerien	t159	98	ST121	-	-	-	x	H2			PEN	
09-01872	244	2009	Deutschland	t284	79	ST121	-	-	-	x	H3		D	PEN, ERY, CLI, CIP, FUS, MFL	Abszess
09-02295	W63	2008	Samoa	t1077	10	ST121	-	-	-	x	H2		A	PEN	SSTI
09-02296	W186	2008	Samoa	t159	12	ST121	-	-	-	x	H2			PEN, ERY	SSTI
09-02298	T12	2008	Tonga	t159	2	ST121	-	-	-	x	H2		A	PEN, ERY	
09-02299	M26	2008	Neuseeland	t159	5	ST121	-	-	-	x	H2		A	PEN	Abszess
09-02301	A48	2008	Neuseeland	t1077	7	ST121	-	-	-	x	H2		A	PEN	Abszess
09-02302	T14	2008	Tonga	t159	11	ST121	-	-	-	x	H2		A	PEN, ERY	
09-02303	W55	2008	Samoa	t1077	10	ST121	-	-	-	x	H2		A	PEN	SSTI
09-02305	W81	2008	Samoa	t1077	8	ST121	-	-	-	x	H2		A	PEN	SSTI
09-02307	F66	2008	Fidschi	t159	6	ST121	-	-	-	x	H2		A	PEN	
09-02308	T8	2008	Tonga	t645	9	ST121	-	-	-	x	H2		A	PEN	
09-02309	T13	2008	Tonga	t159	4	ST121	-	-	-	x	H2		A	PEN	
09-02310	F68	2008	Fidschi	t5448	1	ST121	-	-	-	x	H2		A	PEN	
09-02483	V10504	2008	Deutschland	t284	77	ST121	-	-	-	x	H3		D	PEN	Wundinfektion
09-02512	2007 S263	2007	Belgium	t645	24	ST121	-	-	-	-				CIP1	
09-02513	2008 S207	2008	Belgium	t435	85	ST121	-	-	-	x	H3		D	PEN	Abszess
09-02514	2008 S303	2008	Belgium	t159	46	ST121	-	x	x	-			F	PEN	Impetigo
09-02515	2008 S327	2008	Belgium	t159	64	ST121	-	x	x	-			F	PEN	SSSS
10-00254	544/95	1995	Uganda	t159	89	ST121	-	-	-	x	H2			sensibel	Pyomyositis
10-00255	547/95	1995	Uganda	t645	89	ST121	-	-	-	x	H2			PEN	Pyomyositis
10-00256	549/95	1995	Uganda	t645	101	ST121	-	-	-	x	H4			PEN, TET	Pyomyositis
10-00257	550/95	1995	Uganda	t314	107	ST121	-	-	-	x	H2		B	SXT	Pyomyositis
10-00258	551/95	1995	Uganda	t314	107	ST121	-	-	-	x	H2		B	PEN, ERY, TET, SXT	Pyomyositis
10-00259	553/95	1995	Uganda	t314	105	ST121	-	-	-	x	H2		B	SXT	Pyomyositis
10-00260	556/95	1995	Uganda	NT	100	ST121	-	-	-	x	H2		B	PEN, TET	Pyomyositis
10-00261	600/95	1995	Uganda	t314	100	ST121	-	-	-	x	H2		B	SXT	Pyomyositis
10-00380	PS71 §	1953	UK	t645	17	ST121	-	-	-	x	H2			sensibel	
10-01114	1	2004	Russland	t435	71	ST121	-	-	-	x	H3		D	PEN	
10-01115	2	2004	Südafrika	t317	112	ST121	-	-	-	x	H2		B	PEN	
10-01116	3	2004	Südafrika	t6712	112	ST121	-	-	-	x	H2		B	PEN	
10-01117	4	2004	Russland	t435	71	ST121	-	-	-	x	H2		D	PEN	
10-01119	6	2004	Russland	t435	71	ST121	-	-	-	x	H3		D	PEN, ERY, TET	
10-01120	7	2004	Russland	t308	87	ST121	-	-	-	x	H3		D	PEN	
10-01121	8	2004	Russland	t284	71	ST121	-	-	-	x	H3		D	PEN	
10-01122	9	2004	Russland	t435	71	ST121	-	-	-	x	H3		D	PEN	
10-01123	10	2004	Russland	t2392	71	ST121	-	-	-	x	H3		D	PEN	

Tabelle A.1: CC121-Isolate

Isolat	NRZ-ID	Jahr	Land	spo- type	Haplotyp	MLST	SCCmec	etb	eta	luks/F	PVL	Genotyp	Klade	Resistenz <sup>1</sup>	Infektion
10-01124	11	2004	Russland	t435	71	ST121	—	—	—	x	H3	D	D	PEN	
10-01125	12	2004	Indien	t3204	14	ST121	—	—	—	x	H2	H2		PEN	
10-01126	13	2004	Südafrika	t317	111	ST121	—	—	—	x	H2	H2	B	PEN	
10-01127	14	2004	Südafrika	t4198	110	ST121	—	—	—	x	H2	H2	B	PEN	
10-01128	16	2004	Südafrika	t645	22	ST121	—	—	—	x	H2	H2		PEN	
10-01131	19	2004	Indien	t3204	15	ST121	—	—	—	x	H2	H2		PEN	
10-01132	20	2004	USA	t741	47	ST121	—	x	—	—			F	PEN	
10-01133	21	2004	Peru	t645	31	ST121	—	x	x	—			C	PEN,TET	
10-01134	22	2004	USA	t159	11	ST121	—	—	—	x	H2		A	PEN,ERY	
10-01135	23	2004	USA	t2155	48	ST121	—	x	—	—			F	PEN	
10-01136	24	2004	Südafrika	t314	48	ST121	—	—	—	x	H2	H2	B	PEN	
10-01137	25	2004	Südafrika	t317	113	ST121	—	—	—	x	H2	H2	B	PEN	
10-01140	28	2004	Südafrika	t645	23	ST121	—	—	—	x	H2	H2		PEN	
10-01142	30	2004	Südafrika	t2391	70	ST121	—	—	—	x	H3			sensibel	
10-01146	34	2004	Russland	t284	88	ST121	—	—	—	x	H3		D	PEN	
10-01147	35	2004	USA	t3454	72	ST121	—	—	—	—			D	PEN,ERY	
10-01148	09/754	2009	Tschechien	t659	42	ST123	—	x	x	—			F	PEN	
10-01149	09/573	2009	Tschechien	t159	52	ST121	—	x	x	—			F	PEN,ERY	
10-01150	09/178	2009	Tschechien	t272	66	ST121	—	x	x	—			F	PEN	
10-01151	08/797	2008	Tschechien	t916	39	ST121	—	x	x	—			F	PEN	
10-01152	07/280	2007	Tschechien	t645	63	ST121	—	—	x	—			F	PEN,MUP	
10-01153	05/598	2005	Tschechien	t171	41	ST123	—	x	x	—			F	PEN,FUS	
10-01154	03/459	2003	Tschechien	t169	69	ST121	—	x	—	—			F	PEN	
10-01735	ST20070671	2006	Algerien	t3650	90	ST121	—	—	—	x	H3		F	PEN	
10-01736	ST20081338	2007	Algerien	t170	53	ST121	—	—	—	—			F	PEN	
10-01737	ST20090980	2009	Algerien	t272	18	ST121	—	—	—	x	H2		F	PEN	
10-01739	A870449	1987	Frankreich	t314	109	ST121	—	—	—	x	H2		B	PEN	
10-01740	A890511	1989	Frankreich	t314	116	ST121	—	—	—	x	H2		B	PEN	
10-01741	A910759	1991	Frankreich	t159	33	ST121	—	—	—	x	H3		C	PEN	
10-01742	A970388	1997	Frankreich	t645	32	ST121	—	—	—	x	H2		C	PEN,ERY	
10-01743	A980257	1998	Frankreich	t314	108	ST121	—	—	—	x	H2		B	ERY	
10-01744	HT20030729	2003	Frankreich	t159	91	ST121	—	—	—	x	H2		E	PEN,ERY,CLI,CIP	
10-01745	HT20040111	2004	Frankreich	t3635	26	ST121	—	—	—	x	H2			PEN,SXTi	
10-01746	HT20040449	2004	Frankreich	t159	62	ST121	—	—	—	—			F	PEN,ERY,TET	
10-01747	HT20041082	2004	Frankreich	t1114	83	ST121	—	—	—	x	H3		D	PEN	
10-01748	ST20101201	2009	Indien	t159	25	ST121	—	—	—	x	H2			PEN	
10-01749	ST20101205	2009	Indien	t3204	13	ST121	—	—	—	x	H2			PEN	

Tabelle A.1: CC121-Isolate

Isolat	NRZ-ID	Jahr	Land	spo- type	Haplotyp	MLST	SCCmec	etb	eta	luks/F	PVL	Genotyp	Klade	Resistenz <sup>1</sup>	Infektion
10-01750	HT20050331	2005	Bahamas	t940	36	ST121	—	—	—	x	H2		C	PEN	
10-01751	ST20082025	2008	Bahamas	t2155	28	ST121	—	—	—	—			C	PEN,FUS	
10-01752	ST20080520	2008	Frankreich	t645	20	ST121	—	—	—	x	H2			PEN,TET	
10-01753-1	ST20080533	2008	Frankreich	t159	99	ST121	—	—	—	—				PEN	
10-01753-2	ST20080533	2008	Frankreich	t2155	28	ST121	—	—	—	—			C	PEN,FUS	
10-01754	HT20050874	2005	Paraguay	t159	16	ST121	—	—	—	x	R2			PEN	
10-01755	HT20051161	2005	Paraguay	t645	27	ST121	—	x	x	—			C	PEN	
10-01756	HT20020323	2002	Frankreich	t159	59	ST121	—	—	—	—			F	PEN,ERY	
10-01757	ST20080599	2008	Frankreich	t645	30	ST121	—	—	—	—			C	PEN	
10-01758	HT20020235	2002	Spanien	t284	78	ST121	—	—	—	x	H3		D	PEN,TET	
10-01759	ST20090069	2005	Spanien	t645	38	ST121	—	—	—	x	H2		C	PEN,ERY,TET	
10-01760	ST20090103	2008	Spanien	t435	76	ST121	—	—	—	x	H3		D	PEN	
10-01761	HT20020826	2002	Togo	t159	97	ST121	—	—	—	x	H2		B	PEN, RAM	
10-01762	HT20040170	2004	Togo	t314	117	ST121	—	—	—	x	H2			PEN	
10-01763	HT20070286	2006	Türkei	t435	81	ST121	—	—	—	x	H3		D	PEN	
10-01764	HT20070227	2007	UK	t1596	21	ST121	—	—	—	x	H2			PEN	
10-01765	ST20070579	2001	USA	t645	29	ST121	—	—	—	x	H2		C	PEN	
10-01766	ST20070581	2002	USA	t940	34	ST121	—	—	—	x	H2		C	PEN,ERY,FUS	
11-02367	CIP52	2008	Kambodscha	t159	92	ST121	SCCmec V	—	—	x	H2		C	PEN,ERY,FUS	
11-02368	COP1413	2008	Kambodscha	t159	92	ST121	SCCmec V	—	—	x	H2		E	PEN, OXA, ERY	
11-02369	COP2043	2008	Kambodscha	t9089	92	ST121	SCCmec V	—	—	x	H2		E	PEN, OXA, ERY, TET	
11-02370	2007-025444	2007	Kambodscha	t159	93	ST121	SCCmec V	—	—	x	R2		E	PEN, OXA, ERY	
11-02371	2006-001658	2006	Kambodscha	t159	92	ST121	SCCmec V	—	—	x	H2		E	PEN, OXA, ERY, TET	
11-02372	2003-017409	2007	Kambodscha	t159	92	ST121	—	—	—	x	H2		E	PEN, OXA, ERY	
11-02376	2006-004580	2006	Kambodscha	t159	94	ST121	—	—	—	x	H5		E	PEN, ERY, TET	
11-02377	2006-007990	2006	Kambodscha	t159	92	ST121	—	—	—	x	H2		E	PEN, ERY, TET	
11-02379	2006-012157	2006	Kambodscha	t1425	96	ST121	—	—	—	x	H2		E	PEN, ERY, TET	
11-02382	2006-019471	2006	Kambodscha	t170	92	ST121	—	—	—	x	R2		E	PEN, ERY, TET	
11-02384	S009	2006	Thailand	t7002	94	ST121	—	—	—	x	R1		E	PEN, TET	
11-02385	S027	2006	Thailand	t159	94	ST121	—	—	—	x	R2		E	PEN, TET	
11-02387	S035	2006	Thailand	t7002	94	ST121	—	—	—	x	R2		E	PEN, TET	
11-02388	S096	2007	Thailand	t159	95	ST121	—	—	—	x	H2		E	PEN, TET	
11-02389	S104	2007	Thailand	t159	94	ST121	—	—	—	x	R2		E	PEN, TET	
11-02390	S126	2007	Thailand	t159	94	ST121	—	—	—	x	R2		E	PEN, TET	
11-02391	S148	2007	Thailand	t9090	94	ST121	—	—	—	x	R2		E	PEN, TET	
11-02392	S176	2007	Thailand	t159	94	ST121	—	—	—	x	R2		E	PEN	
11-02394	S194	2007	Thailand	t159	94	ST121	—	—	—	x	R2		E	PEN, TET	

Tabelle A.1: CC121-Isolate

Isolat	NRZ-ID	Jahr	Land	<i>spa</i> - type	Haplotyp	MLST	SCCmec	<i>etb</i>	<i>eta</i>	<i>lukS/F</i>	PVL	Klade	Resistenz <sup>1</sup>	Infektion
11-02395	S267	2007	Thailand	t159	94	ST121	–	–	–	x	H2	E	PEN,TET	
97-01419	489	1996	Bulgarien	t270	82	ST121	–	–	–	x	H3	D	PEN,TET	Furunkel
97-01451	490	1996	Bulgarien	t308	84	ST121	–	–	–	x	H3	D	PEN	Furunkel
97-01452	-	1997	Bulgarien	t272	102	ST121	–	–	–	x	H3		PEN,TET	Furunkel
97-01548	VA 066336	1997	Deutschland	t645	40	ST121	–	–	x	–		F	PEN,ERY,TET	Wundinfektion
97-01848-1	A 5642	1997	Deutschland	t159	51	ST121	–	–	x	–		F	ERY,TET	
97-01973	VA 4753	1997	Deutschland	t159	51	ST121	–	–	x	–		F	sensibel	Impetigo
98-01618	66-006310	1998	Deutschland	t159	R	ST121	–	–	x	–		F	PEN	Wundinfektion
99-02025	B 7967	1999	Österreich	t645	67	ST121	–	–	–	–		F	PEN,ERY	

\* getestet Antibiotika: PEN, Penicillin G; OXA, Oxacillin; TPL, Teicoplanin; VAN, Vancomycin; GEN, Gentamicin; TET, Tetracyclin; CIP, Ciprofloxacin; MFL, Moxifloxacin; SXT, Trimethoprim/Sulfamethoxazol; PHO, Fosfomycin; FUS, Fusidinsäure; ERY, Erythromycin; CLI, Clindamycin; RAM, Rifampicin; DAP, Daptomycin; MUP, Mupirocin; LNZ, Linezolid and TGC, Tigecycline



Tabelle A.2: Identifizierte Polymorphismen in 154 CC121-Isolaten im Vergleich zum Referenzstamm 96-01618 (CC121)

SNP-ID	Auswirkung der Mutation	Ursprung	Abgeleitet	Transition(si)& Transversion(ver)	Position (N315 Genom)	ORF in N315
au201-1	synonym	G	C	ver	110592	SA0098
au201-2	indel	-	insertion(3nt)		110685-87	SA0098
au201-3	nicht-synonym	G	A	si	110905	SA0098
au201-4	nicht-synonym	C	T	si	110680	SA0098
au202-1	nicht-synonym	C	A	ver	130307	SA0113
au202-2	synonym	G	A	si	130312	SA0113
au203-1	nicht-synonym	C	T	si	142486	SA0123
au203-2	nicht-synonym	C	T	si	142553	SA0123
au205-2	nicht-synonym	T	G	ver	213260	SA0182
au206-1	nicht-synonym	G	A	si	239470	SA0202
au206-2	nicht-synonym	G	T	ver	239713	SA0202
au206-3	nicht-synonym	G	A	si	239833	SA0202
au207-1	nicht-synonym	C	A	ver	263379	SA0220
au207-2	synonym	G	A	si	263213	SA0220
au207-3	nicht-synonym	C	A	ver	263130	SA0220
au207-4	nicht-synonym	A	G	si	263145	SA0220
au208-1	nicht-synonym	G	A	si	310194	SA0256
au208-2	nicht-synonym	G	A	si	310173	SA0256
au208-3	nicht-synonym	G	A	si	310227	SA0256
au209-1	nicht-synonym	C	T	si	355330	SA0299
au209-2	synonym	A	G	si	355430	SA0299
au210-1	nicht-synonym	C	T	si	369435	SA0311
au210-2	nicht-synonym	A	G	si	369642	SA0311
au210-3	nicht-synonym	G	A	si	369610	SA0311
au210-4	synonym	G,T,A	A		369491	SA0311
au210-5	nicht-synonym	G	A	si	369726	SA0311
au210-6	synonym	A	G	si	369671	SA0311
au212-1	synonym	T	C	si	406083	SA0346
au212-2	synonym	C	T	si	405930	SA0346
au212-3	nicht-synonym	G	A	si	406142	SA0346
au212-4	nicht-synonym	C	G	ver	405865	SA0346
au212-5	nicht-synonym	A	C	ver	406102	SA0346
au213-1	nicht-synonym	A	G	si	435049	SA0376
au214-1	synonym	C	A	ver	480909	SA0419
au214-2	nicht-synonym	C	T	si	481086	SA0419
au214-3	nicht-synonym	C	T	si	480946	SA0419
au215-1	nicht-synonym	T	C	si	496066	SA0431
au215-2	nicht-synonym	C	T	si	495953	SA0431
au218-1	synonym	T	C	si	592482	SA0507
au218-2	nicht-synonym	G	C	ver	592619	SA0507
au218-3	nicht-synonym	G	A	si	592548	SA0507
au218-4	nicht-synonym	A	C	ver	592604	SA0507
au219-1	nicht-synonym	C	T	si	599756	SA0512
au219-2	nicht-synonym	C	T	si	599615	SA0512
au219-3	nicht-synonym	A	C	ver	599609	SA0512
au219-4	synonym	G	A	si	599502	SA0512
au222-1	synonym	T	C	si	795280	SA0697
au223-1	nicht-synonym	G	A	si	836822	SA0730
au224-1	nicht-synonym	G	A	si	897017	SA0791
au224-2	nicht-synonym	G	C	ver	897311	SA0791
au224-3	nicht-synonym	C	T	si	897011	SA0791
au225-1	nicht-synonym	G	T	ver	974479	SA0859
au227-1	synonym	T	C	si	1181583	SA1044
au227-2	synonym	T	C	si	1181427	SA1044
au227-3	synonym	A	C	ver	1181565	SA1044
au228-1	synonym	A	G	si	1265225	SA1115
au230-1	synonym	A	T	ver	1401407	SA1226
au230-2	synonym	G	A	si	1401386	SA1226
au230-3	nicht-synonym	T	G	ver	1401703	SA1226
au231-1	synonym	T	G	ver	1514387	SA1308
au231-2	nicht-synonym	G	T	ver	1514242	SA1308
au232-1	synonym	T	C	si	1617607	SA1412
au232-2	nicht-synonym	T	C	si	1617715	SA1412

Tabelle A.2: Identifizierte Polymorphismen in 154 CC121-Isolaten im Vergleich zum Referenzstamm 96-01618 (CC121)

SNP-ID	Auswirkung der Mutation	Ursprung	Abgeleitet	Transition(si)& Transversion(ver)	Position (N315 Genom)	ORF in N315
au232-3	nicht-synonym	G	A	si	1617669	SA1412
au232-4	nicht-synonym	G	T	ver	1617699	SA1412
au232-5	nicht-synonym	G	C	ver	1617890	SA1412
au233-1	nicht-synonym	C	T	si	1704949	SA1496
au233-2	synonym	T	C	si	1704911	SA1496
au233-3	synonym	A	G	si	1705007	SA1496
au234-1	nicht-synonym	C	T	si	1766929	SA1545
au234-2	nicht-synonym	C	T	si	1767092	SA1545
au234-3	nicht-synonym	T	A	ver	1767034	SA1545
au234-4	nicht-synonym	G	A	si	1767061	SA1545
au234-6	nicht-synonym	A	G	si	1767164	SA1545
au236-1	nicht-synonym	T	C	si	2054252	SA1814
au236-2	synonym	T	A	ver	2054010	SA1814
au236-3	nicht-synonym	A	T	ver	2053948	SA1814
au236-4	synonym	A	G	si	2054112	SA1814
au236-5	nicht-synonym	A	T	ver	2054130	SA1814
au237-1	synonym	T	C	si	2173422	SA1924
au237-2	synonym	A	C	ver	2173584	SA1924
au237-3	synonym	G	A	si	2173379	SA1924
au237-4	nicht-synonym	C	T	si	2173448	SA1924
au237-5	synonym	G	A	si	2173560	SA1924
au237-6	nicht-synonym	G	T	ver	2173562	SA1924
au238-1	synonym	C	A	ver	2263830	SA1991
au238-2	nicht-synonym	C	T	si	2263826	SA1991
au238-3	nicht-synonym	G	T	ver	2263764	SA1991
au238-4	nicht-synonym	C	A	ver	2263645	SA1991
au238-5	synonym	C	T	si	2263623	SA1991
au238-6	nicht-synonym	G	T	ver	2263541	SA1991
au239-1	indel	-	insertion(3nt)		2384349-51	SA2120
au240-1	nicht-synonym	G	A	si	2485731	SA2214
au240-2	nicht-synonym	A	G	si	2485737	SA2214
au242-1	synonym	T	C	ver	2701455	SA2410
au242-2	synonym	G	A	si	2701653	SA2410
au243-1	nicht-synonym	A	G	si	2782074	SA2470
au243-2	nicht-synonym	C	T	si	2782090	SA2470
au243-3	nicht-synonym	T	C	si	2781984	SA2470
au243-4	indel	T	deletion(1nt)		2782109	SA2470
au243-5	nicht-synonym	T	C	si	2782116	SA2470
au243-6	synonym	A	G	si	2782046	SA2470
au243-7	nicht-synonym	T	C	si	2782290	SA2470
au244-1	nicht-synonym	G	A	si	207004	SA0177
au244-2	synonym	T	C	si	207150	SA0177
au244-3	nicht-synonym	T	G	ver	207349	SA0177
au244-4	synonym	C	T	si	207069	SA0177
au244-5	nicht-synonym	C	T	si	207199	SA0177
au245-1	synonym	C	T	si	291850	SA0240
au245-2	indel	A	deletion(1nt)		291482	SA0240
au245-3	nicht-synonym	T	G	ver	291729	SA0240
au247-1	nicht-synonym	A	G	si	418908	SA0361
au247-2	nicht-synonym	G	A	si	418853	SA0361
au247-3	nicht-synonym	G	A	si	418878	SA0361
au247-4	synonym	T	C	si	418717	SA0361
au247-5	synonym	C	T	si	418816	SA0361
au247-6	nicht-synonym	A	T	ver	418893	SA0361
au250-1	nicht-synonym	T	C	si	665916	SA0572
au250-2	nicht-synonym	G	A	si	665737	SA0572
au251-1	nicht-synonym	G	A	si	764669	SA0671
au251-2	nicht-synonym	G	A	si	764659	SA0671
au251-3	indel	T	deletion(1nt)		764595	SA0671
au252-1	nicht-synonym	C	T	si	926406	SA0821
au252-2	synonym	G	A	si	926061	SA0821
au252-3	nicht-synonym	C	T	si	926453	SA0821
au253-1	nicht-synonym	C	T	si	1017687	SA0896

Tabelle A.2: Identifizierte Polymorphismen in 154 CC121-Isolaten im Vergleich zum Referenzstamm 96-01618 (CC121)

SNP-ID	Auswirkung der Mutation	Ursprung	Abgeleitet	Transition(si)& Transversion(ver)	Position (N315 Genom)	ORF in N315
au254-1	nicht-synonym	A	T	ver	1049696	SA0924
au254-2	nicht-synonym	T	C	si	1049646	SA0924
au254-3	nicht-synonym	C	T	si	1049672	SA0924
au254-4	synonym	A	G	si	1049726	SA0924
au254-5	nicht-synonym	C	T	si	1049793	SA0924
au254-6	synonym	C	T	si	1049620	SA0924
au255-1	nicht-synonym	G	A	si	1232894	SA1089
au255-2	synonym	A	G	si	1232797	SA1089
au256-1	synonym	C	A	ver	1307743	SA1150
au256-2	nicht-synonym	C	T	si	1307948	SA1150
au257-1	nicht-synonym	C	T	si	1575100	SA1366
au257-2	nicht-synonym	T	C	si	1575082	SA1366
au258-1	nicht-synonym	A	G	si	1693668	SA1487
au258-2	nicht-synonym	T	C	si	1693526	SA1487
au259-1	synonym	C	T	si	1739473	SA1524
au260-1	nicht-synonym	T	G	ver	1781591	SA1554
au260-2	indel	T	deletion(1nt)		1781592	SA1554
au260-3	nicht-synonym	C	T	si	1781535	SA1554
au260-4	nicht-synonym	G	A	si	1781303	SA1554
au260-5	nicht-synonym	T	C	si	1781659	SA1554
au260-6	nicht-synonym	C	T	si	1781372	SA1554
au260-7	nicht-synonym	C	G	ver	1781440	SA1554
au261-1	nicht-synonym	C	A	ver	1824318	SA1585
au261-2	nicht-synonym	C	T	si	1824093	SA1585
au261-3	nicht-synonym	G	A	si	1824094	SA1585
au262-1	synonym	T	C	si	1906188	SA1669
au267-1	nicht-synonym	C	A	ver	2332560	SA2071
au268-1	synonym	C	T	si	2359745	SA2099
au268-2	nicht-synonym	A	G	si	2359357	SA2099
au269-1	nicht-synonym	G	A	si	2517330	SA2240
au269-2	nicht-synonym	G	A	si	2517255	SA2240
au270-1	nicht-synonym	C	T	si	2537906	SA2260
au270-2	synonym	T,G	A		2537775	SA2260
au270-3	nicht-synonym	C	T	si	2537950	SA2260
au272-1	nicht-synonym	C	G	ver	2638773	SA2348
au272-2	nicht-synonym	C	T	si	2638792	SA2348
au272-3	nicht-synonym	C	T	si	2638669	SA2348
au272-4	nicht-synonym	C	T	si	2638753	SA2348
au272-5	nicht-synonym	G	A	si	2638791	SA2348
au273-1	nicht-synonym	G	T	ver	2680915	SA2393
au273-2	nicht-synonym	T	C	si	2681042	SA2393
au274-1	nicht-synonym	T	C	si	1845230	SA2490
au274-2	synonym	G	A	si	1845127	SA2490
au274-3	nicht-synonym	A	C	ver	1845248	SA2490
au275-1	nicht-synonym	C	T	si	1099509	SA0969
au275-2	synonym	C	G	ver	1099813	SA0969
au277-1	nicht-synonym	C	T	ver	2743500	SA2439
au277-2	nicht-synonym	C	T	ver	2743275	SA2439
au277-3	nicht-synonym	G	A	si	2743223	SA2439
au277-4	synonym	A	C	ver	2743384	SA2439
au277-5	nicht-synonym	G	A	si	2743220	SA2439
au279-1	nicht-synonym	C	T	si	847385	SA0740
au279-2	nicht-synonym	C	T	si	847357	SA0740
au279-3	nicht-synonym	C	T	si	847154	SA0740
au279-4	nicht-synonym	C	T	si	847209	SA0740
au279-5	nicht-synonym	G	C	ver	847412	SA0740
au280-1	nicht-synonym	G	A	si	921290	SA0817
au280-2	synonym	A	C	ver	921282	SA0817
au280-3	nicht-synonym	C	T	si	921158	SA0817
au281-1	indel	-	deletion(12nt)		2753300-311	SA2445
au281-2	synonym	A	G	si	2753365	SA2445
au281-3	nicht-synonym	A	G	si	2753480	SA2445
au281-4	nicht-synonym	G	A	si	2753484	SA2445

Tabelle A.2: Identifizierte Polymorphismen in 154 CC121-Isolaten im Vergleich zum Referenzstamm 96-01618 (CC121)

SNP-ID	Auswirkung der Mutation	Ursprung	Abgeleitet	Transition(si)& Transversion(ver)	Position (N315 Genom)	ORF in N315
au282-2	nicht-synonym	G	A	si	2385822	SA2121
au282-3	nicht-synonym	G	A	si	2385997	SA2121
au284-1	synonym	G	A	si	21549	SA0015
au284-2	synonym	A	G	si	21876	SA0015
au284-3	synonym	G	T	ver	21849	SA0015
au285-1	nicht-synonym	G	A	si	139091	SA0119
au286-1	nicht-synonym	C	T	si	267666	SA0223
au286-2	nicht-synonym	G	A	si	267642	SA0223
au287-1	synonym	A	G	si	477514	SA0416
au287-2	synonym	T	C	si	477727	SA0416
au287-3	nicht-synonym	G	A	si	477636	SA0416
au288-1	nicht-synonym	C	A	ver	638859	SA0545
au289-1	nicht-synonym	G	T	ver	750827	SA0656
au289-2	nicht-synonym	C	G	ver	750755	SA0656
au290-1	nicht-synonym	G	C	ver	957089	SA0843
au293-1	nicht-synonym	T	C	si	1084881	SA0958
au293-2	synonym	G	A	si	1085206	SA0958
au294-1	nicht-synonym	A	C	ver	1274398	SA1121
au294-2	nicht-synonym	G	A	si	1274294	SA1121
au294-3	nicht-synonym	T	C	si	1274355	SA1121
au295-1	nicht-synonym	G	A	si	1325103	SA1163
au295-2	nicht-synonym	T	C	si	1324997	SA1163
au296-1	synonym	G	A	si	1373290	SA1199
au296-4	nicht-synonym	G	A	si	1373325	SA1199
au298-1	nicht-synonym	G	A	si	1559595	SA1349
au298-2	nicht-synonym	G	C	ver	1559430	SA1349
au299-1	indel	-	insertion(3nt)		1582583-85	SA1377
au299-2	indel	-	deletion(9nt)		1582559-67	SA1377
au299-3	nicht-synonym	G	A	si	1582368	SA1377
au300-1	nicht-synonym	C	T	si	1887159	SA1652
au300-2	synonym	T	C	si	1886980	SA1652
au300-3	synonym	C	T	si	1886824	SA1652
au301-1	nicht-synonym	C	T	si	1932719	SA1685
au301-2	synonym	G	T	ver	1932733	SA1685
au301-3	indel	T	insertion(1nt)		1932728	SA1685
au302-1	nicht-synonym	A	T	ver	1974165	SA1724
au303-1	synonym	A	G	si	2001591	SA1749
au303-2	nicht-synonym	G	T	ver	2001610	SA1749
au304-1	synonym	G	A	si	2097960	SA1858
au304-2	nicht-synonym	G	T	ver	2098060	SA1858
au304-3	synonym	T	C	si	2098038	SA1858
au306-1	nicht-synonym	C	T	si	2216849	SA1963
au307-1	nicht-synonym	A	G	si	2272461	SA2001
au308-1	nicht-synonym	C	A	ver	2405135	SA2140
au309-1	synonym	A	G	si	2459905	SA2188
au309-2	nicht-synonym	A	G	si	2460149	SA2188
au319-1	nicht-synonym	G	A	si	402863	SA0344
au319-7	synonym	G	A	si	402780	SA0344
au319-8	synonym	C	A	ver	402930	SA0344
au325-1	nicht-synonym	A	T	ver	598728	SA0511
au326-1	synonym	A,G	A	si	627855	SA0533
au326-2	nicht-synonym	-	insertion(1nt)		627841	SA0533
au326-3	synonym	C	T	si	627795	SA0533
au332-1	synonym	G	A	si	883327	SA0776
au332-2	nicht-synonym	C	T	si	883391	SA0776
au332-3	synonym	G	A	si	883225	SA0776
au332-4	synonym	C	T	si	883285	SA0776
au335-1	nicht-synonym	C	T	si	950534	SA0837
au335-2	nicht-synonym	G	C	ver	950769	SA0837
au335-3	indel	-	insertion(3nt)		950489	SA0837
au336-1	nicht-synonym	C	G	ver	1042844	SA0917
au336-2	nicht-synonym	C	T	si	1042538	SA0917
au336-3	nicht-synonym	A	T	ver	1042480	SA0917

Tabelle A.2: Identifizierte Polymorphismen in 154 CC121-Isolaten im Vergleich zum Referenzstamm 96-01618 (CC121)

SNP-ID	Auswirkung der Mutation	Ursprung	Abgeleitet	Transition(si)& Transversion(ver)	Position (N315 Genom)	ORF in N315
au341-1	nicht-synonym	C	T	si	1275916	SA1122
au341-2	synonym	A	C	ver	1275996	SA1122
au341-3	nicht-synonym	G	C	ver	1275945	SA1122
au346-1	nicht-synonym	G	A	si	1369964	SA1197
au346-2	nicht-synonym	G	T	ver	1370133	SA1197
au346-3	nicht-synonym	G	A	si	1370213	SA1197
au346-4	nicht-synonym	C	G	ver	1369884	SA1197
au346-5	nicht-synonym	G	A	si	1370081	SA1197
au346-6	nicht-synonym	A	G	si	1369934	SA1197
au350-1	nicht-synonym	C	A	ver	1511921	SA1306
au350-2	nicht-synonym	C	A	ver	1511938	SA1306
au350-3	nicht-synonym	G	A	si	1511763	SA1306
au350-4	nicht-synonym	G	A	si	1511763	SA1306
au352-1	nicht-synonym	G	T	ver	1517161	SA1310
au352-2	nicht-synonym	G	A	si	1516834	SA1310
au352-3	nicht-synonym	C	G	ver	1517145	SA1310
au352-4	nicht-synonym	G	A	si	1517050	SA1310
au352-5	nicht-synonym	C	T	si	1516928	SA1310
au355-1	nicht-synonym	C	T	si	1556430	SA1346
au355-2	synonym	C	T	si	1556474	SA1346
au358-1	nicht-synonym	C	T	si	1563091	SA1352
au358-2	nicht-synonym	G	A	si	1563332	SA1352
au366-1	synonym	G	A	si	1701849	SA1493
au366-2	synonym	A	G	si	1701762	SA1493
au366-3	nicht-synonym	T	A	ver	1701959	SA1493
au371-1	synonym	A	T	ver	1737417	SA1522
au371-2	synonym	T	C	si	1737528	SA1522
au382-1	nicht-synonym	T	A	ver	1843706	SA1608
au382-2	nicht-synonym	C	T	si	1843571	SA1608
au383-1	synonym	T	C	si	1848942	SA1614
au383-2	nicht-synonym	G	A	si	1848806	SA1614
au383-3	nicht-synonym	C	A	ver	1848750	SA1614
au383-4	nicht-synonym	T	A	ver	1848799	SA1614
au384-1	nicht-synonym	T	G	ver	1849487	SA1615
au384-2	nicht-synonym	T	A	ver	1849791	SA1615
au384-3	synonym	T	C	si	1849558	SA1615
au384-4	synonym	T	C	si	1849693	SA1615
au391-1	synonym	C	T	si	2162848	SA1913
au391-2	synonym	C	T	si	2162755	SA1913
au391-3	nicht-synonym	T	C	si	2162544	SA1913
au391-4	nicht-synonym	G	A	si	2162888	SA1913
au392-1	nicht-synonym	T	C	si	2267501	SA1995
au392-2	nicht-synonym	C	T	si	2267834	SA1995
au392-3	nicht-synonym	C	A	ver	2267756	SA1995
au392-4	nicht-synonym	G	C	ver	2267801	SA1995
au392-5	indel	TAT	deletion(3nt)		2267617-19	SA1995
au392-6	nicht-synonym	A	G	si	2267850	SA1995
au392-7	nicht-synonym	C	A	ver	2267823	SA1995
au392-8	nicht-synonym	G	T	ver	2267781	SA1995
au395-1	nicht-synonym	G	A	ver	2344729	SA2084
au395-2	nicht-synonym	C	T	si	2344750	SA2084
au395-3	synonym	C	T	si	2344944	SA2084
au395-4	nicht-synonym	C	T	si	2344733	SA2084
au395-5	synonym	C	T	si	2344872	SA2084

Tabelle A.3: Primer für die Verifizierung von SNPs in den CC121-Isolaten.

Locus	ORF	Annotation	Primer-F	Primer-R	Position (N315 Genom)	Produktgröße (Bp)	dHPLC T/(°C)
au200	SA0016	adenylosuccinate synthase	TTGCTTTCGAAGGTAATTGG	CGTGCATTTTCAGGTAATTCTTT	23216	23628	56.3
au201	SA0098	aminoacylase	TGAAGATGAACGGCGAAAT	CCTGTTTTTCATTTGTGCTCAT	110529	110948	57.3
au202	SA0113	ornithine cyclodeaminase	AAACGCGTGTGTGTTACGA	TTTTGTGTTGTGCTGTTGA	130244	130731	56.3
au203	SA0123	UDP-glucose 4-epimerase	TGGTTTTATTGGGTGCGATT	AATTAAATGTCGTCCGTTGG	142173	142618	54.7
au205	SA0182	indole-3-pyruvate decarboxylase	TTAGTTGCAATGAGCCATCA	TTCAGCAACAGCAGGATTTT	213236	213690	54.3
au206	SA0202	gamma-glutamyltranspeptidase	TC TTGCGGCTAAAAATGGTAA	TGGCGCGAATATTTTTCAGT	239439	239855	56.3
au207	SA0220	glycerophosphodiester phosphodiesterase	TGAAAAAGGCGCATGTTC'TTC	TCCATTTTGAAGCATTTTCA	263008	263471	54.6
au208	SA0256	6-phospho-beta-glucosidase	OCGCAAAATCAATTTGAAGGT	GCGCAAAATGCACAAAATAA	310060	310488	55.6
au209	SA0299	carbohydrate kinase, PfkB family	TTGCAGAAACATCAAAATCCTG	TCATTTTGTGGGAAGAAACG	355123	355563	56.0
au210	SA0311	trimethylamine dehydrogenase	TGGAATTTCCACACCATCA	CCGTAAACGCATCAACAAA	369364	369789	57.7
au212	SA0346	cystathionine beta-lyase	TGACAGCACCCAGCAATTAAA	ATTGCTTTTGTCTTTTGCAT	405743	406170	56.2
au213	SA0376	GMP synthase	AAGTTGAACGTGCCAATGAA	TTGATCCCAATTTGCTTTATG	434800	435239	57.2
au214	SA0419	cystathionine gamma-synthase	CAACATGTGGCTTTGCGTTTA	ATGCCACTGTGTAATTTGA	480802	481293	56.5
au215	SA0431	NADH-glutamate synthase	AAGGGGCGATTACTTTTCAT	AACGCATGTTTGAACACCA	495821	496294	55.8
au218	SA0507	N-acyl-L-amino acid amidohydrolase	TCAAATTTGGGATTATGGTGTG	CAAACTCATGAACCAATTTGA	592338	592835	54.9
au219	SA0512	aminotransferase homologue	AAATTTACCTGCTGCACAA	TGCAACTTTTGCAAAAGCTTA	599411	599841	56.6
au220	SA0537	phosphomethylpyrimidine kinase	AAGCTTTGCCACCTTTAATGA	TGCAAGCAGATTTGAAAACG	631634	632116	56.3
au222	SA0697	glycerate kinase	CAGTTGCAAGCCAAATTGAA	TGCACTTGAAAATCCGACA	795048	795547	55.6
au223	SA0730	2,3-diphosphoglycerate mutase	CGGAAATTTTTCGGAACAGA	TTGATTTGCTTTTGATTTGTCG	836703	837148	55.3
au224	SA0791	glycerate dehydrogenase	TGACGTTGTTATGTGGCAAAA	TTAAACCCCTTGCAACCTTCTT	896974	897414	56.8
au225	SA0859	thimet oligopeptidase homologue	TACTGAATTCGGCAAGAAATG	ATGCAACTTCAGCGACAAAA	974341	974741	55.4
au226	SA0945	dihydrolipoamide S-acetyltransferase	AAAGAAATCGCAGCTGAACAA	TTTTGAGCAATACGGCCAAT	1073758	1074161	56.2
au227	SA1044	dihydroorotase	ATGAAGGGAACGCAGTAAA	CAATCCGAATGGTGTCTTTT	1181330	1181731	56.4
au228	SA1115	riboflavin kinase / FAD synthase	CAATGGCATTCGGATTTTTC	CATTGCTTTTGTGCAACTCA	1264780	1265262	53.8
au229	SA1177	transketolase	TTGGTTCACCGCAATAAAGCA	TGGATTTGTTTGAACCCAGCA	1339372	1339785	54.8
au230	SA1226	aspartate semialdehyde dehydrogenase	AAAAAGCTGGTGCAATCGTT	TGCTTTCTTTTGTCAAGCGTTA	1401265	1401747	56.1
au231	SA1308	30S ribosomal protein S1	ATGCACCAAAAGTTTGGCCAAT	TTCCGGCTTTCAC'TAATTTCAA	1513993	1514478	54.8
au232	SA1412	oxygen-independent coproporphyrinogen oxidase III	TTCAITGGCCATCTAATGCAA	CAITTTGAAGCAAAATCCTGATGA	1617553	1618016	54.6
au233	SA1496	glutamyl-tRNA reductase	ATTTGCCCTTGCACCAATAAT	CGAATTTGCCCATGAAGATTTT	1704604	1705096	56.1
au234	SA1545	D-3-phosphoglycerate dehydrogenase	TGCAGTGAATGCACCTAAAA	TAACAATGCGCCAGTTTTC	1766721	1767216	54.2
au236	SA1814	succinyl-L-diaminopimelate desuccinylase	ATGTTGCACCCCATGTTTCAAA	CATCGAATTTGATGTGCCATT	2053853	2054311	55.6

Tabelle A.3: Primer für die Verifizierung von SNPs in den CC121-Isolaten.

Locus	ORF	Annotation	Primer-F	Primer-R	Position (N315 Genom)	Produktgröße (Bp)	dHPLC T/(°C)
au237	SA1924	aldehyde dehydrogenase	GCAAAATGCTGCTGCTAAATTTT	TTAATGGCGGAATGGGTAGAAA	2173216	2173675	56.7
au238	SA1991	6-phospho-beta-galactosidase	TTTGTGTTGGCAATGCGTGTA	ATTTCAATTGTCATGGTCACG	2263407	2263853	55.1
au239	SA2120	amino acid amidohydrolase	GCAATTGCTTTTATGCGTCAAT	TCAAAACAGGGGCAATTACA	2383969	2384377	57.3
au240	SA2214	adenosylmethionine-8-amino-7-oxononanoate aminotransferase	ACCCCTGCGATTTCATCATTT	TAAAAATTGCCCATTTCTACGC	2485384	2485811	55.7
au242	SA2410	anaerobic ribonucleoside-triphosphate reductase	ACTCGTGGCGCAAAATTTCTTTT	ACGGATCGTTTTTTGTGCTGTTT	2701426	2701886	56.5
au243	SA2470	histidinol dehydrogenase	TGTTTGAATCGACGCATGTT	TTGCTGCAATTGACTTATGGAA	2781905	2782326	57.2
au244	SA0177	<i>argJ</i> , arginine biosynthesis bifunctional protein homologue	TGGCAAAATCACCAAGTTGAA	TGCGTTAATAACGCACATAATCA	206898	207388	57.0
au245	SA0240	sorbitol dehydrogenase homologue	TTGAAAGTGGGAATGAAGTT	CAAAATTTCTGTGCCCATTTGA	291411	291896	57.9
au247	SA0361	truncated phosphoglycerate mutase	CATGGCGAATCAAAATCGAA	CCATCAAAACAACGAATGACAA	418661	419144	55.2
au248	SA0527	glucosamine-6-phosphate isomerase	AAGAGATGGCAITTTAATCAGCA	AATGGCTTGCTTTGGAACAT	622181	622659	54.6
au250	SA0572	esterase/lipase	TTTAAACGATCACGTTTGAGGA	AAAGGATCATGTTCCCATTT	665507	665947	54.2
au251	SA0671	urea amidolyase	GAAATTTGTGCCATCAGAAATCA	TTTAAAGCGTGCAGAAATCTGA	764197	764696	54.2
au252	SA0821	<i>argH</i> , argininosuccinate lyase	TGCATTTTTCAACTGGCTCAT	TTTGAAGTGGTTTCAAGCAA	926012	926505	54.8
au253	SA0896	<i>menD</i> , menaquinone biosynthesis protein	GGTTAGAAAAAATGGCAATGCT	GCCAAACAACCGTTCAAAAT	1017529	1017968	54.1
au254	SA0924	<i>purN</i> , phosphoribosylglycinamide formyltransferase	CGATTTTGTGCATCAGGTTCA	TTTTGAATCGTCCGGTCTAA	1049523	1050001	54.8
au255	SA1089	<i>sucD</i> , succinyl-CoA synthetase	AAAAATGAAACTGGGGCAAC	CCGATTTTCACCCAATCATAACAA	1232627	1233078	56.9
au256	SA1150	<i>gluA</i> , glutamine-ammonia ligase	TGAAAAAGGGGAACCAACTT	AATCCGCGTGCATTTTAAAG	1307593	1308024	56.1
au257	SA1366	glycine dehydrogenase subunit 1	TGCGATTTGCAAGCAAGAAATA	CAATTGAACCGTAAAAGTTTGG	1575017	1575507	55.6
au258	SA1487	<i>folC</i> , folylpolyglutamate synthase	TTTACGTCCCGTTTATTTGA	CAATTGGCTTTGCATGTTGT	1693267	1693720	55.8
au259	SA1524	malate dehydrogenase homologue	CGCAATTTTGTGAAGGAGGTT	CTTTAGCAACATTTACGAGCAA	1739355	1739845	56.2
au260	SA1554	<i>acsA</i> , acetyl-CoA synthetase	ATFTGTGGCCGCTTAITTTGA	TTTGTAGCGCCATTTAAACCA	1781281	1781729	55.8
au261	SA1585	proline dehydrogenase homologue	TTTAGGGGAATTTGTCCGTA	GAAATTGCGTGGGTTTAAACA	1824023	1824497	54.7
au262	SA1669	<i>cifG</i> , fumarate hydratase, class-II	ACACATTTTACAAGATGCAACG	TGCAACCATTGTTAACATTTCA	1905900	1906352	57.2
au263	SA1735	manganese-dependent inorganic pyrophosphatase	CCTGAAATTGCGCGGTTTAAT	TTTTTCGAGAAACAACACCAGA	1987761	1988244	55.1
au265	SA1965	<i>glmM</i> , femD, phosphoglucosamine-mutase	ACAAAGCGCTTGAACAAGAA	TTGTGCAAAATCTTTCAGCATC	2225913	2226346	55.4
au267	SA2071	<i>moeB</i> , molybdopterin biosynthesis protein moeB	TGCGTTTGATTATTGGTATGG	TTTAAATGCTGGCAATTGTGG	2332305	2332760	56.0
au268	SA2099	monooxygenase	ATCCGGATTGTGCAAAAAGAA	TGGGCAATTTTACCGATTTT	2359332	2359800	54.7
au269	SA2240	para-nitrobenzyl esterase chain A	TGCACGCAATAAAGCACAAAT	TGCAAAATTCAGCAAGCCATA	2517061	2517511	55.0

Tabelle A.3: Primer für die Verifizierung von SNPs in den CC121-Isolaten.

Locus	ORF	Annotation	Primer-F	Primer-R	Position (N315 Genom)	Produktgröße (Bp)	dHPLC T/(°C)
au270	SA2260	glucose 1-dehydrogenase	TGCGTGGGACAAATTTTAATG	TTCCACGAATCATCACTTAAC	2537713	2538172	57.6
au271	SA2327	pyruvate oxidase	ATGTAGGATGGCCAAAATGGA	TCATCCGGCAAGATTGTTTT	2612925	2613400	54.9
au272	SA2348	<i>ertN</i> , squalene desaturase	TTAATGCCAGATTTTTCACCT	TTTTCGCATGATACGTTTGCT	2638403	2638864	54.3
au273	SA2393	2-dehydropanoate 2-reductase	AGACACGCCAAAACAATAACA	ATGCATGATTGCAACTGTCT	2680864	2681340	55.7
au274	SA2490	<i>pkcA</i> , phosphoenolpyruvate carboxykinase	TTTCAAACTTTTCGACGACACAA	TGCTTTAAAAATGTGTTGCAGA	1844849	1845280	54.4
au275	SA0969	glycerophosphoryl diester phosphodiesterase	CGCACCACAAAATTATGTTTGA	TTTCGGGATTGTTTGTTTGAA	1099466	1099873	55.6
au277	SA2439	<i>sasF</i> , conserved hypothetical protein (Cooper + Feil)	AATGCCTCAACGCAACATTA	CAACGATTTTGGCGATTTGAT	2743068	2743554	54.9
au279	SA0740	hypothetical protein (Cooper + Feil)	TTGGAACATTTTATTATGGCTTT	ACGGTCGATTTTGACCTTTTA	847062	847504	54.0
au280	SA0817	NADH-dependent flavin oxidoreductase (Cooper + Feil)	TGAAGAAGTGCTTAAAGCGAAA	AATTTTCATCTTCCCGTCCAT	921075	921477	57.0
au281	SA2445	hypothetical protein (Cooper + Feil)	TTAAAAATGGCGGATGATGT	TTTCGCTGGAATATGATTGG	2753158	2753616	55.4
au282	SA2121	<i>hutI</i> , imidazolonepropionase (Cooper + Feil)	CCCATTTAAACATTTGGTGGA	GCTTCTTTTGGCTTTTGTGCAT	2385678	2386152	56.0
au284	SA0015	<i>dnaC</i> , replicative DNA helicase (Cooper + Feil)	AAAGTTGCAACGCATGAAGA	TTGCTCAATCGAAACCAGATT	21445	21939	56.8
au285	SA0119	diaminopimelate	TTGAATGTGGCGCTTTATT	TCCTCCTTTGTTTGGCTTAAA	138812	139214	56.3
au286	SA0223	UDP-glucose 4-epimerase	TGCAATCGTTTGTGAGAAGCAA	TGGTCGGTTCAAAGATGAAA	267449	267872	57.7
au287	SA0416	carboxylesterase	ATTAAAAATTGGCGCAGCACT	GCATTGAAAAAGCGAATAACA	477346	477781	55.0
au288	SA0545	phosphotransacetylase	GCGCAAAAACCTTGATCTTGA	TGCCAAAAGCTTAATGCTGATT	638508	638919	56.4
au289	SA0656	probable N-acetylglucosamine-6-phosphate deacetylase	ATTGTTGCAATTGGAGAAAGG	ACGTTTCAAAGCGCTTCTTTT	750368	750863	54.9
au290	SA0843	3-oxoacyl- synthase	TATGGCAATTGCGAGGTTTCA	TTGCTTCAATTCCACCTGTT	956696	957156	57.8
au291	SA0881	nucleotidase	GGCATCAAAAACGTTGGATTGA	GCGAAAATGGATGTGTTGA	1001452	1001936	54.6
au292	SA0902	HisC homolog	TGGAACGATTGTAAATGTGTT	TCCTTCGCTGAATTGGAAGA	1023317	1023721	54.8
au293	SA0958	myo-inositol-1(or 4)-monophosphatase	TTTATGGATTTGCCCAAGGA	TTCAATGGTTTCAAGGGTTG	1084827	1085247	54.8
au294	SA1121	processing proteinase	AAGAAAATGGCCAAACCGATAA	TTGCCGACAAACATAAACAGAA	1273973	1274433	53.2
au295	SA1163	aspartate kinase homolog	AACAAAAATACCCGCTTCCTT	AAAGTTTCCAAATTCGGTGGT	1324852	1325298	53.9
au296	SA1199	anthranilate synthase	AAATCCAAACCATTCATGCCCTA	TTTCGTGCATCTCTTTTGGAT	1373094	1373518	55.3
au298	SA1349	dihydrolipoamide dehydrogenase	TTGATCCAAACATGTGCCAAT	TGGTGTATCGGAATGGAAT	1559232	1559653	54.6
au299	SA1377	glucokinase	TCAAAATTGAGCAGGTGCAAA	GCGCAGAAAATAGGTCATTT	1582274	1582676	54.7
au300	SA1652	uroporphyrinogen decarboxylase	AAGCAACCATGTTGCTTCA	TTTCGCAACCAGAAATATCGAA	1886778	1887220	55.3
au301	SA1685	A/G-specific adenine	AATTGCCACATACCATGCAA	TCCCCAAAAGATCCTGATCAATTT	1932656	1933123	54.6



Tabelle A.3: Primer für die Verifizierung von SNPs in den CC121-Isolaten.

Locus	ORF	Annotation	Primer-F	Primer-R	Position (N315 Genom)	Produktgröße (Bp)	dHPLC T/(°C)
au302	SA1724	adenylosuccinate lyase	ACGGAAAGGTGTTTTCGTGTTT	AAATTTGCCGTTGAAATTCG	1974092 1974559	428	55.6
au303	SA1749	aspartate transaminase	GCCCAAAGCTTAAATGAACAA	ATGTTTGAAGGTTTGACACCA	2001475 2001949	434	55.4
au304	SA1858	dihydroxy-acid dehydratase	TTGGTTTGACGGGTATTTT	ATGCATCATCAATTGCTTCG	2097921 2098372	412	57.5
au305	SA1874	alanine racemase	TCGCATTGATGACCAATTTACA	CTTGTTATTTGAAGGCGTGTTT	2122020 2122457	396	55.8
au306	SA1963	mannitol-1-phosphate dehydrogenase	TTGCAGACGTCAATGAAGAAA	TCCTTTTTCACAAACCCATTTCG	2216805 2217249	403	55.2
au307	SA2001	oxidoreductase	TTCTTCTTGTGATGGAATGGA	AAACAGCAGAAAGCGGTTAAA	2272199 2272623	383	55.1
au308	SA2140	esterase	TGCCATTAAATTGGCGATATG	TTCCATTGGAGATTGTTGGA	2404832 2405246	375	55.8
au309	SA2188	nitrite reductase	TGCATTTCCATCAACCAATTC	TAACCCGCATCATGCTTTTCAA	2459852 2460274	383	55.7
au319	SA0344	homocysteine methyltransferase	TTGCCCATACATTAGGACCA	ACGTTAACGCTCACCCCTGTC	402524 402944	381	54.5
au325	SA0511	UDP-glucose 4-epimerase related protein	CGATTTCAGCAACCTACGACA	TAGGATCAAAATCCCCATTTCG	598347 598835	449	55.6
au326	SA0533	CoA-ligase	TATGAAATCACCGGTGCAAA	CGACCGGAACTTAAAGCAAA	627507 627914	368	56.1
au332	SA0776	aminotransferase NifS homologue	AAATGCTTTGGACCAACAGG	CCCAGCTCTAACCCGCTACAC	883076 883486	371	56.7
au335	SA0837	2-isopropylmalate synthase	AAGGGATGGTATGCAACAAAG	TCCTACAGTGTCCGCAAATG	950317 950794	437	52.7
au336	SA0917	phosphoribosylaminoimidazole carboxylase	CCAAAAGTTGTCCCGTTTCATT	TGGTGCAATCTCGTTTCACAT	1042431 1042882	412	54.2
au341	SA1122	proteinase	CGTGTGTGATATTGCCGGTAG	GTAGCGCTTGTACAGATTGA	1275589 1276070	442	54.8
au346	SA1197	prephenate dehydrogenase	CACCTCAATGCTTGCATACCG	AACTTGCAGCTGGTGGTTT	1369809 1370243	395	55.2
au350	SA1306	glycerol-3-phosphate dehydrogenase	TTTGTGATGATGCAGCAACTG	TGTTCTTTCAGAAAATGACAA	1511681 1512098	377	55.3
au352	SA1310	probable L-asparaginase	ATGAAAAGGCCCGCTCATAAA	AAAACCTTGTGTGCGAGTT	1516729 1517186	418	56.7
au355	SA1346	dehydrogenase E2	TGTTATTTGTGCTGTTTTTCACT	TTCCACGATATCAGGAACAA	1556201 1556621	380	56.0
au358	SA1352	geranyltransferase homologue	TTTATCTTCTGCGCCATCTTT	TCGACGAGGAAATTAACAAA	1563060 1563555	454	54.8
au366	SA1493	uroporphyrinogen III synthase	TGCATCGATTGTGTGATGTTG	CACAAACAAATGACATGCAAAAG	1701528 1701985	415	52.8
au371	SA1522	acetyl-CoA carboxylase	TGTGCAACAAACGCTGATTT	AAATTTTGGTATGGCGCATC	1737361 1737849	449	55.9
au382	SA1608	S-adenosylmethionine synthetase	CGTTTGTTTCAAATCAAGCAT	ACATCAAAGCGCATGTCATT	1843323 1843768	405	57.5
au383	SA1614	o-succinylbenzoic acid synthetase	ACAAACGATGTTGCCCTTTTCA	AATTGAGCAATGGTTGGAAGA	1848623 1849098	435	56.5
au384	SA1615	O-succinylbenzoic acid-CoA ligase	TGCAATCAATTGTGCCCTTTG	TTTAAACAGCAACACCCGGAAA	1849417 1849846	390	55.4
au391	SA1913	UDP-GlcNAc 2-epimerase	TTAAATGATGGCGCTTCTTC	ATGCTGCATCGCATTTGTTA	2162492 2162915	384	55.2
au392	SA1995	tagatose-6-phosphate kinase	TTTCATCAAAGTGGCTGGTTA	TGGTTTATTTGGCGGTGAAT	2267457 2267887	391	55.4
au395	SA2084	urease alpha subunit	ATATGCAATTOGCAAGATTCA	ACCAITTTGCATCGCCATTTA	2344580 2344983	364	56.3

Tabelle A.4: Primer für die Verifizierung der Toxingene *eta*, *etb* und *lukS/F-PV* und zum Nachweis der *attB* des Sa2-Phagen

Gen	Primer	Sequenz (5'→3')	Richtung	Produkt- größe (bp)	Annealing- Temperatur
<i>eta</i>	eta f	CTA GTG CAT TTG TTA TTC AA	vorwärts	119	45°C
	eta r	TGC ATT GAC ACC ATA GTA CT	rückwärts		
<i>etb</i>	etb f	ACG GCT ATA TAC ATT CAA TT	vorwärts	200	45°C
	etb r	TCC ATC GAT AAT ATA CCT AA	rückwärts		
<i>lukS/F-PV</i>	<i>lukS/F-PV</i> f	ATC ATT AGG TAA AAT GTC TGG ACA TGA TCC A	vorwärts	432	55°C
	<i>lukS/F-PV</i> r	GCA TCA AST GTA TTG GAT AGC AAA AGC	rückwärts		
<i>attB</i> (Sa2)	attB_PVLint-L f	TGG GAA TAA ATG TCG ACG AA	vorwärts	220	58°C
	attB_PVLint-R r	CGT GCA TGA GCT TGA GTG TT	rückwärts		

## Ergebnisse der BaTS-Analysen

Tabelle A.5: Ergebnisse der BaTS-Analyse zur Phylogenie/*spa*-Typ Assoziation

<i>spa</i> -Typ	MC-Wert	<i>p-value</i>
t645	5,01	0,01
t314	4,53	0,01
t284	1,30	1
t940	3,08	0,01
t435	1,38	1
t170	1	1
t272	1	1
t2155	1	1
<b>AI</b> =7,07; p=0		
<b>PS</b> =51,13; p=0		

Tabelle A.6: Ergebnisse der BaTS-Analyse zur Phylogenie/Geographie Assoziation

Region	MC-Wert	<i>p-value</i>
Afrika	5,08	0,01
Asien	3,34	0,01
Europa	9,08	0,01
Mittelamerika	10,45	0,01
Nordamerika	2,00	0,05
Südamerika	1,17	1,00
Ozeanien	3,00	0,01
<b>AI</b> =2,56; p=0		
<b>PS</b> =25,72; p=0		

Tabelle A.7: BaTS-Werte für die Assoziation zwischen Phylogenie und klinischem Phänotyp

Infektionstyp	MC-Wert	<i>p-value</i>
oberflächlich	13,00	0,01
tiefgehend	13,83	0,01
<b>AI</b> =2,69E-06; p=0		
<b>PS</b> =1,00; p=0		

Tabelle A.8: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Virulenzgen-Gehalt

<b>Virulenzgen</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
<i>lukS/F-PV</i>	35,98	0,01
<i>eta/etb</i>	5,07	0,01
<b>AI</b> =1,80; p=0		
<b>PS</b> =11,84; p=0		

Tabelle A.9: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Erythromycin

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Erythromycin	3,10	0,07
<b>AI</b> =4,24;		
p=0,29		
<b>PS</b> =23,50;		
p=0,15		

Tabelle A.10: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Penicillin

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Penicillin	13,13	0,45
<b>AI</b> =1,18; p=0		
<b>PS</b> =11,81; p=0		

Tabelle A.11: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Tetracyclin

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Tetracyclin	2,50	0,44
<b>AI</b> =3,32; p=0		
<b>PS</b> =18,73; p=0		

Tabelle A.12: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Trimethoprim/Sulfamethoxazol

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Trimethoprim- /Sulfamethoxazol	2,00	0,03
<b>AI</b> =0,58; p=0,02		
<b>PS</b> =5,00; p=0,02		

Tabelle A.13: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Clindamycin

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Clindamycin	1,00	1,00
<b>AI</b> =0,63; p=0,61		
<b>PS</b> =3,00; p=1,00		

Tabelle A.14: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Ciprofloxacin

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Ciprofloxacin	1,00	1,00
<b>AI</b> =0,37; p=0,09		
<b>PS</b> =3,00; p=1,00		

Tabelle A.15: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Oxacillin

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Oxacillin	1,46	1,00
<b>AI</b> =0,25; p=0,39		
<b>PS</b> =1,55; p=1,00		

Tabelle A.16: BaTS-Analyse zeigt die Assoziation zwischen Phylogenie und Fusidinsäure

<b>Antibiotika</b>	<b>MC-Wert</b>	<b><i>p-value</i></b>
Fusidinsäure	2,25	0,10
<b>AI</b> =0,90; p=0		
<b>PS</b> =6,00 p=0		

## ABKÜRZUNGSVERZEICHNIS

### A-E

A.bidest	zweifach destilliertes Wasser
Abb.	Abbildung
ATP	Adenosintriphosphat
APS	Adenosin-5'-phosphosulfat
bp	Basenpaare
BS	Bootstrap-Werte
caMRSA	engl. „ <i>community acquired MRSA</i> “
CC	klonaler Komplex; engl. „ <i>Clonal Complex</i> “
CHIPS	Chemotaxis inhibiting Protein
dN	Anzahl der nicht-synonymen Substitutionen
DNA	Desoxyribonukleinsäure
dHPLC	denaturierende Hochdruckchromatographie
dNTP	Desoxyribonukleosidtriphosphate
ddNTP	Didesoxyribonukleosidtriphosphate
dS	Anzahl der synonymen Substitutionen
Dsg-1	Desmoglein-1
Dsg-3	Desmoglein-3
EDTA	Ethylendiamintetraessigsäure
engl.	englisch
ESS	engl. „ <i>Effective Sample Size</i> “
ETA	Exfoliativ-Toxin A
ETB	Exfoliativ-Toxin B
<i>et al.</i>	<i>et alii</i> (weiter Autoren)

### F-J

h	Stunde, Stunden
haMRSA	engl. „ <i>hospital acquired MRSA</i> “
HCL	Chlorwasserstoff
HI	Homoplasie-Index
<i>IS</i>	Insertionselement

### K-O

kb	Kilobasenpaare ( $1 \times 10^3$ )
KI	Konfidenzintervall
laMRSA	engl. „ <i>livestock associated MRSA</i> “
lat.	latein
M	Molar
$\mu g$	Mikrogramm
mg	Milligramm
min.	Minute, Minuten

μl	Mikroliter
ml	Milliliter
mM	Millimolar
ng	Nanogramm
nm	Nanometer
Mb	Megabasenpaare (1 x 10 <sup>6</sup> )
MCMC	engl. „ <i>Markov Chain Monte Carlo</i> “
MCMCMC	engl. „ <i>Metropolis-coupled Markov Chain Monte Carlo</i> “
MGE	Mobile genetische Elemente
MHK	Minimale Hemmkonzentration
ML	<i>Maximum-Likelihood</i>
MLST	Multi-Lokus-Sequenz-Typisierung
MRSA	Methicillin-Resistente <i>Staphylococcus aureus</i>
MSSA	Methicillin-Sensible <i>Staphylococcus aureus</i>
MST	<i>Minimum-Spanning-Tree</i>
n.a.	Information nicht verfügbar; engl. „ <i>not available</i> “
NGS	engl. „ <i>Next-Generation Sequencing</i> “
NRZ	Nationales Referenzzentrum
ns	nicht-synonym
OD	Optische Dichte
ORF	engl. „Open Reading Frame“
<b>P-T</b>	
PBP	Penicillin-Bindeprotein
PCR	Polymerase-Kettenreaktion; engl. „ <i>Polymerase Chain Reaction</i> “
PFGE	Pulsfeld-Gel-Elektrophorese
pH	lat. <i>potentium hydrogenii</i>
PP	engl. „ <i>Posterior Probability</i> “
PML	Polymorphkernige Leukozyten
PP <sub>i</sub>	anorganisches Pyrophosphat
PVL	Panton-Valentine Leukozidin
RNA	Ribonukleinsäure
rpm	Rotationen pro Minute
s	synonym
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
<i>Saint</i>	Integrasegruppe der <i>S. aureus</i> Prophagen
SaPI <sub>n</sub>	Pathogenitätsinsel
SCC <sub>mec</sub>	Staphylococcal Cassette Chromosome <i>mec</i>
SCIN	<i>S. aureus</i> Komplement Inhibitor
SLV	Einzelnukeotidvarianten; eng. „ <i>Single Locus Variants</i> “
SNP	Einzelnukeotidpolymorphismus; engl. „ <i>Single Nucleotide Polymorphism</i> “
SSSS	<i>Staphylococcal Scalded Skin Syndrom</i>

SSTI	<i>skin and soft tissue infections</i>
ST	Sequenztyp; engl. „Sequence Type“
TBE	Tris-Borat-EDTA
TEAA	Triethylammoniumacetat
TMRCA	engl. „ <i>time to most recent common ancestor</i> “
Tris	Tris-(hydroxymethyl)-aminomethan
TSB	Tryptic Soy Broth
TSST-1	Toxische-Schock-Syndrom-Toxin-1
TSS	Toxische-Schock-Syndrom
<i>Tn</i>	Transposon
si	Transition
ver	Transversion
<b>V-Z</b>	
XML	engl. „ <i>Extensible Markup Language</i> “
z.B.	zum Beispiel

## Gennamen

<i>agr</i>	Accessory Gene Regulator
<i>arcA-D</i>	Gencluster zum Argininabbau
<i>bbp</i>	Bindet Bone-Sialoprotein
<i>blaZ</i>	$\beta$ -Lactamase Gen für Penizillinresistenz
<i>cadDX</i>	Cadmium Resistenz Operon
<i>ccrA</i>	Rekombinase-Gen A ( <i>cassette chromosome recombinase A</i> )
<i>ccrB</i>	Rekombinase-Gen B
<i>ccrC</i>	Rekombinase-Gen C
<i>chp</i>	Chemotaxis-Inhibitor protein CHIPS
<i>clfA/B</i>	Clumping Factor A und B
<i>cna</i>	Kollogen-Adhäsion
<i>coa</i>	Koagulase
<i>ear</i>	vermutlich $\beta$ -Laktamase Protein
<i>ebps</i>	Elastin-bindende Protein
<i>ecb</i>	Extrazelluläres Komplement-bindendes Protein
<i>edinA/B</i>	epidermal cell differentiation inhibitor A/B
<i>edin-C</i>	Epidermaler Zelldifferenzierung Inhibitor C
<i>egc</i>	Enterotoxin-Cluster
<i>ermC</i>	Erythromycin-Resistenz
<i>eta</i>	Exfoliatives Toxin A
<i>etb</i>	Exfoliatives Toxin B
<i>fnbA/B</i>	Fibronectin (Fn)-bindende Proteine
<i>fusB</i>	Fusidinsäure-Resistenz Protein FusB



<i>guaA</i>	Guanosin-monophosphat-(GMP)-Synthase
<i>hsdR/M/S</i>	Komponenten des RM-Systems
<i>sdrC/D/H</i>	Adhäsine (Serine-Aspartate-repeat containing proteins)
<i>hly</i>	$\beta$ -Hämolysin
<i>icaA/B/C/D</i>	Operon für Syntheseapparat des interzellulären Polysaccharid Adhäsins (PIA)
<i>isdA/B/C/D/E</i>	Eisen-regulierte oberflächen-determinierende Proteine A-E
<i>lpl</i>	Lipoprotein Gen Cluster
<i>lukDE</i>	Komponenten des Leukozidin DE Toxins
<i>lukS/F-PV</i>	Komponenten des Panton-Valentine Leukozidin S und F
<i>mecA</i>	Methicillin-Resistenzprotein 2a
<i>orfX</i>	Integration von SCC <i>mec</i>
<i>rpsR</i>	ribosomales Protein der 30S-Untereinheit
<i>sak</i>	Staphylokinase
<i>sar</i>	Staphylococcal Accessory Regulator
<i>sasG</i>	Oberflächenprotein G
<i>sbi</i>	Staphylokokken-Immunglobulin-bindendes Protein
<i>scn</i>	Staphylokokken Komplement Inhibitor SCIN
<i>sea</i>	Enterotoxin A
<i>seb</i>	Staphylokokken Enterotoxin B
<i>sec</i>	Enterotoxin C
<i>sed</i>	Enterotoxin D
<i>seh</i>	Enterotoxin H
<i>sej</i>	Enterotoxin J
<i>sek</i>	Enterotoxin K
<i>sel</i>	Enterotoxin L
<i>sep</i>	Enterotoxin P
<i>seq</i>	Enterotoxin Q
<i>ser</i>	Enterotoxin R
<i>set</i>	Staphylokokken Exotoxine
<i>sspA/B/P</i>	Staphopain A/B/P
<i>splA/B/E</i>	Staphylokokken Serinproteasen A/B/E
<i>spl</i>	Staphylokokken Serinprotease
<i>tetK</i>	Tetracyclin-Resistenz
<i>tst</i>	Toxin-1 des Toxin-Schock-Syndroms

# Literaturverzeichnis

- [1] M. Achtman, G. Morelli, P. Zhu, T. Wirth, I. Diehl, B. Kusecek, A.J. Vogler, D.M. Wagner, C.J. Allender, W.R. Easterday, et al. Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17837, 2004.
- [2] M. Aires de Sousa and H. Lencastre. Bridges from hospitals to the laboratory: genetic portraits of methicillin-resistant *Staphylococcus aureus* clones. *FEMS Immunology & Medical Microbiology*, 40(2):101–111, 2006.
- [3] D. Alland, T.S. Whittam, M.B. Murray, M.D. Cave, M.H. Hazbon, K. Dix, M. Kokoris, A. Duesterhoeft, J.A. Eisen, C.M. Fraser, et al. Modeling bacterial evolution with comparative-genome-based marker systems: application to *Mycobacterium tuberculosis* evolution and pathogenesis. *Journal of Bacteriology*, 185(11):3392–3399, 2003.
- [4] M. Amagai, P.J. Koch, T. Nishikawa, and J.R. Stanley. Pemphigus vulgaris antigen (desmoglein 3) is localized in the lower epidermis, the site of blister formation in patients. *Journal of Investigative Dermatology*, 106(2):351–355, 1996.
- [5] M. Amagai, T. Yamaguchi, Y. Hanakawa, K. Nishifuji, M. Sugai, and J.R. Stanley. Staphylococcal exfoliative toxin b specifically cleaves desmoglein 1. *Journal of Investigative Dermatology*, 118(5):845–850, 2002.
- [6] T. Baba, F. Takeuchi, M. Kuroda, H. Yuzawa, K. Aoki, A. Oguchi, Y. Nagai, N. Iwama, K. Asano, T. Naimi, et al. Genome and virulence determinants of high virulence community-acquired mrsa. *The Lancet*, 359(9320):1819–1827, 2002.
- [7] L. Baker, T. Brown, M.C. Maiden, and F. Drobniewski. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. *Emerging Infectious Diseases*, 10(9):1568, 2004.
- [8] M. Barber and M. Rozwadowska-Dowzenko. Infection by penicillin-resistant staphylococci. *Lancet*, 2(6530):641, 1948.

- [9] G. Blum, M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschäpe, and J. Hacker. Excision of large dna regions termed pathogenicity islands from trna-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infection and Immunity*, 62(2):606–614, 1994.
- [10] S. Boyle-Vavra and R.S. Daum. Community-acquired methicillin-resistant *Staphylococcus aureus*: the role of panton–valentine leukocidin. *Laboratory Investigation*, 87(1):3–9, 2006.
- [11] S. Breurec, C. Fall, R. Pouillot, P. Boisier, S. Brisse, F. Diene-Sarr, S. Djibo, J. Etienne, MC Fonkoua, JD Perrier-Gros-Claude, et al. Epidemiology of methicillin-susceptible *Staphylococcus aureus* lineages in five major african towns: high prevalence of panton-valentine leukocidin genes. *Clinical Microbiology and Infection*, 17(4):633–639, 2011.
- [12] S. Breurec, SB Zriouil, C. Fall, P. Boisier, S. Brisse, S. Djibo, J. Etienne, MC Fonkoua, JD Perrier-Gros-Claude, R. Pouillot, et al. Epidemiology of methicillin-resistant *Staphylococcus aureus* lineages in five major african towns: emergence and spread of atypical clones. *Clinical Microbiology and Infection*, 17(2):160–165, 2011.
- [13] M.L. Brown, F.P. O’Hara, N.M. Close, R.M. Mera, L.A. Miller, J.A. Suaya, and H. Amrine-Madsen. Prevalence and sequence variation of panton-valentine leukocidin in methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* strains in the united states. *Journal of Clinical Microbiology*, 50(1):86–90, 2012.
- [14] M. Bukowski, B. Wladyka, and G. Dubin. Exfoliative toxins of *Staphylococcus aureus*. *Toxins*, 2(5):1148–1165, 2010.
- [15] H.F. Chambers. The changing epidemiology of *Staphylococcus aureus*? *Emerging Infectious Diseases*, 7(2):178, 2001.
- [16] K. Chheng, S. Tarquinio, V. Wuthiekanun, L. Sin, J. Thaipadungpanit, P. Amornchai, N. Chanpheaktra, S. Tumapa, H. Putschhat, N.P.J. Day, et al. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* associated with pediatric infection in cambodia. *PLoS One*, 4(8):e6630, 2009.
- [17] M.A. Chlebowicz, K. Nganou, S. Kozytska, J.P. Arends, S. Engelmann, H. Grundmann, K. Ohlsen, J.M. Van Dijl, and G. Buist. Recombination between *ccrC* genes in a type v (5c2&5) staphylococcal cassette chromosome *mec* (*sccmec*) of *Staphylococcus aureus* st398 leads to conversion from methicillin resistance to methicillin susceptibility *In vivo*. *Antimicrobial Agents and Chemotherapy*, 54(2):783–791, 2010.

- [18] P. Chongtrakool, T. Ito, X.X. Ma, Y. Kondo, S. Trakulsomboon, C. Tiensasitorn, M. Jamklang, T. Chavalit, J.H. Song, and K. Hiramatsu. Staphylococcal cassette chromosome *mec* (*sccmec*) typing of methicillin-resistant *Staphylococcus aureus* strains isolated in 11 asian countries: a proposal for a new nomenclature for *sccmec* elements. *Antimicrobial Agents and Chemotherapy*, 50(3):1001–1012, 2006.
- [19] A.E. Clatworthy, E. Pierson, and D.T. Hung. Targeting virulence: a new paradigm for antimicrobial therapy. *Nature Chemical Biology*, 3(9):541–548, 2007.
- [20] D.C. Coleman, J.P. Arbuthnott, H.M. Pomeroy, and T.H. Birkbeck. Cloning and expression in *Escherichia coli* and *Staphylococcus aureus* of the beta-lysin determinant from *Staphylococcus aureus*: evidence that bacteriophage conversion of beta-lysin activity is caused by insertional inactivation of the beta-lysin determinant. *Microbial Pathogenesis*, 1(6):549–564, 1986.
- [21] T. Conceição, M. Aires-de Sousa, N. Pona, MJ Brito, C. Barradas, R. Coelho, T. Sardinha, L. Sancho, G. de Sousa, M. do Céu Machado, et al. High prevalence of st121 in community-associated methicillin-susceptible *Staphylococcus aureus* lineages responsible for skin and soft tissue infections in portuguese children. *European Journal of Clinical Microbiology & Infectious Diseases*, 30(2):293–297, 2011.
- [22] B. Cribier, G. Prevost, P. Couppie, V. Finck-Barbancon, E. Grosshans, and Y. Piemont. *Staphylococcus aureus* leukocidin: A new virulence factor in cutaneous infections? *Dermatology*, 185(3):175–180, 1992.
- [23] M.I. Crisóstomo, H. Westh, A. Tomasz, M. Chung, D.C. Oliveira, and H. De Lencastre. The evolution of methicillin resistance in *Staphylococcus aureus*: similarity of genetic backgrounds in historically early methicillin-susceptible and-resistant isolates and contemporary epidemic clones. *Proceedings of the National Academy of Sciences*, 98(17):9865–9870, 2001.
- [24] S.J. Dancer, R. Garratt, J. Saldanha, H. Jhoti, and R. Evans. The epidermolytic toxins are serine proteases. *FEBS letters*, 268(1):129–132, 1990.
- [25] M.A. de Sousa, T. Conceicao, C. Simas, and H. De Lencastre. Comparison of genetic backgrounds of methicillin-resistant and-susceptible *Staphylococcus aureus* isolates from portuguese hospitals and the community. *Journal of Clinical Microbiology*, 43(10):5150–5157, 2005.
- [26] M.A. De Sousa, I.S. Sanches, ML Ferro, MJ Vaz, Z. Saraiva, T. Tendeiro, J. Serra, and H. De Lencastre. Intercontinental spread of a multidrug-resistant

- methicillin-resistant *Staphylococcus aureus* clone. *Journal of Clinical Microbiology*, 36(9):2590–2596, 1998.
- [27] RH Deurenberg, PS Beisser, MJ Visschers, C. Driessen, and EE Stobberingh. Molecular typing of methicillin-susceptible *Staphylococcus aureus* isolates collected in the yogyakarta area in indonesia, 2006. *Clinical Microbiology and Infection*, 16(1):92–94, 2010.
  - [28] R.H. Deurenberg and E.E. Stobberingh. The evolution of *Staphylococcus aureus*. *Infection, Genetics and Evolution*, 8(6):747–763, 2008.
  - [29] RH Deurenberg, C. Vink, S. Kalenic, AW Friedrich, CA Bruggeman, and EE Stobberingh. The molecular evolution of methicillin-resistant *Staphylococcus aureus*. *Clinical microbiology and infection*, 13(3):222–235, 2006.
  - [30] U. Dobrindt, B. Hochhut, U. Hentschel, and J. Hacker. Genomic islands in pathogenic and environmental microorganisms. *Nature Reviews Microbiology*, 2(5):414–424, 2004.
  - [31] Janina Dordel. *Untersuchungen zur Mikroevolution von Methicillin-resistenten Staphylococcus aureus im klonalen Komplex CC5*. Dissertation, Technische Universität Braunschweig, 2012.
  - [32] A. Drummond and A. Rambaut. Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1):214, 2007.
  - [33] M.C. Enright, N.P.J. Day, C.E. Davies, S.J. Peacock, and B.G. Spratt. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *Journal of Clinical Microbiology*, 38(3):1008–1015, 2000.
  - [34] M.C. Enright, D.A. Robinson, G. Randle, E.J. Feil, H. Grundmann, and B.G. Spratt. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (mrsa). *Proceedings of the National Academy of Sciences*, 99(11):7687–7692, 2002.
  - [35] B. Ewing and P. Green. Base-calling of automated sequencer traces usingphred. ii. error probabilities. *Genome Research*, 8(3):186–194, 1998.
  - [36] B. Ewing, L.D. Hillier, M.C. Wendl, and P. Green. Base-calling of automated sequencer traces usingphred. i. accuracy assessment. *Genome Research*, 8(3):175–185, 1998.
  - [37] D. Falush and R. Bowden. Genome-wide association mapping in bacteria? *Trends in Microbiology*, 14(8):353–355, 2006.

- [38] J. Fan, M. Shu, G. Zhang, W. Zhou, Y. Jiang, Y. Zhu, G. Chen, S.J. Peacock, C. Wan, W. Pan, et al. Biogeography and virulence of *Staphylococcus aureus*. *PLoS One*, 4(7):e6216, 2009.
- [39] E.J. Feil, J.E. Cooper, H. Grundmann, D.A. Robinson, M.C. Enright, T. Berendt, S.J. Peacock, J.M. Smith, M. Murphy, B.G. Spratt, et al. How clonal is *Staphylococcus aureus*? *Journal of Bacteriology*, 185(11):3307, 2003.
- [40] W.M. Fitch. Toward defining the course of evolution: minimum change for a specific tree topology. *Systematic Biology*, 20(4):406–416, 1971.
- [41] J.D. Fraser and T. Proft. The bacterial superantigen and superantigen-like proteins. *Immunological Reviews*, 225(1):226–243, 2008.
- [42] JH Freer and JP Arbuthnott. Toxins of *Staphylococcus aureus*. *Pharmacology & Therapeutics*, 19(1):55–106, 1982.
- [43] A. Genestier, M. Michallet, G. Prévost, G. Bellot, L. Chalabreysse, S. Peyrol, F. Thivolet, J. Etienne, G. Lina, F.M. Vallette, et al. *Staphylococcus aureus* panton-valentine leukocidin directly targets mitochondria and induces bax-independent apoptosis of human neutrophils. *Journal of Clinical Investigation*, 115(11):3117, 2005.
- [44] B. Ghebremedhin, MO Olugbosi, AM Raji, F. Layer, RA Bakare, B. König, and W. König. Emergence of a community-associated methicillin-resistant *Staphylococcus aureus* strain with a unique resistance profile in southwest nigeria. *Journal of Clinical Microbiology*, 47(9):2975–2980, 2009.
- [45] W. Göbel, J. Hacker, and U. Dobrindt. *Pathogenomics: Genome Analysis of Pathogenic Microbes*. Wiley-VCH, 2006.
- [46] R.V. Goering, R.M. Shawar, N.E. Scangarella, F.P. O’Hara, H. Amrine-Madsen, J.M. West, M. Dalessandro, J.A. Becker, S.L. Walsh, L.A. Miller, et al. Molecular epidemiology of methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* isolates from global clinical trials. *Journal of Clinical Microbiology*, 46(9):2842–2847, 2008.
- [47] C. Goerke, R. Pantucek, S. Holtfreter, B. Schulte, M. Zink, D. Grumann, B.M. Broker, J. Doskar, and C. Wolz. Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *Journal of Bacteriology*, 191(11):3462, 2009.
- [48] AR Gomes, H. Westh, and H. De Lencastre. Origins and evolution of methicillin-resistant *Staphylococcus aureus* clonal lineages. *Antimicrobial Agents and Chemotherapy*, 50(10):3237–3244, 2006.

- [49] A. Gravet, P. Couppie, O. Meunier, E. Clyti, B. Moreau, R. Pradinaud, H. Monteil, and G. Prevost. *Staphylococcus aureus* isolated in cases of impetigo produces both epidermolysin a or b and luke-lukd in 78% of 131 retrospective and prospective cases. *Journal of Clinical Microbiology*, 39(12):4349–4356, 2001.
- [50] H. Grundmann, S. Hori, M.C. Enright, C. Webster, A. Tami, E.J. Feil, and T. Pitt. Determining the genetic structure of the natural population of *Staphylococcus aureus*: a comparison of multilocus sequence typing with pulsed-field gel electrophoresis, randomly amplified polymorphic dna analysis, and phage typing. *Journal of Clinical Microbiology*, 40(12):4544–4546, 2002.
- [51] G.G. Guay, S.A. Khan, and D.M. Rothstein. The *tetK* gene of plasmid pt181 of *Staphylococcus aureus* encodes an efflux protein that contains 14 transmembrane helices. *Plasmid*, 30(2):163–166, 1993.
- [52] H. Hahn, S.H.E. Kaufmann, and T.F. Schulz. *Medizinische Mikrobiologie und Infektiologie*. Springer DE, 2008.
- [53] D. Harmsen, H. Claus, W. Witte, J. Rothgänger, H. Claus, D. Turnwald, and U. Vogel. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. *Journal of Clinical Microbiology*, 41(12):5442–5448, 2003.
- [54] S.R. Harris, E.J.P. Cartwright, M.E. Török, M.T.G. Holden, N.M. Brown, A.L. Ogilvy-Stuart, M.J. Ellington, M.A. Quail, S.D. Bentley, J. Parkhill, et al. Whole-genome sequencing for analysis of an outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study. *The Lancet Infectious Diseases*, 2012.
- [55] S.R. Harris, E.J. Feil, M.T.G. Holden, M.A. Quail, E.K. Nickerson, N. Chantratita, S. Gardete, A. Tavares, N. Day, J.A. Lindsay, et al. Evolution of mrsa during hospital transmission and intercontinental spread. *Science*, 327(5964):469, 2010.
- [56] A.A. Hartmann. Staphylococci of the normal human skin flora. *Archives of dermatological research*, 261(3):295–302, 1978.
- [57] K. Hiramatsu, L. Cui, M. Kuroda, and T. Ito. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends in Microbiology*, 9(10):486–493, 2001.
- [58] M.T.G. Holden, E.J. Feil, J.A. Lindsay, S.J. Peacock, N.P.J. Day, M.C. Enright, T.J. Foster, C.E. Moore, L. Hurst, R. Atkin, et al. Complete genomes of

- two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26):9786, 2004.
- [59] A. Holmes, M. Ganner, S. McGuane, TL Pitt, BD Cookson, and AM Kearns. *Staphylococcus aureus* isolates carrying panton-valentine leucocidin genes in england and wales: frequency, characterization, and association with clinical disease. *Journal of Clinical Microbiology*, 43(5):2384–2390, 2005.
  - [60] S. Holtfreter, D. Grumann, M. Schmudde, HTT Nguyen, P. Eichler, B. Strommenger, K. Kopron, J. Kolata, S. Giedrys-Kalemba, I. Steinmetz, et al. Clonal distribution of superantigen genes in clinical *Staphylococcus aureus* isolates. *Journal of Clinical Microbiology*, 45(8):2669–2680, 2007.
  - [61] Robert Koch Institut. Auftreten und verbreitung von mrsa in deutschland 2010. *Epidemiologische Bulletin*, 26:233–244, 2011.
  - [62] T. Ito, Y. Katayama, K. Asada, N. Mori, K. Tsutsumimoto, C. Tiensasitorn, and K. Hiramatsu. Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 45(5):1323, 2001.
  - [63] M.P. Jevons. “celbenin”-resistant staphylococci. *British Medical Journal*, 1(5219):124–125, 1961.
  - [64] J. Kaneko and Y. Kamio. Bacterial two-component and hetero-heptameric pore-forming cytolytic toxins: structures, pore-forming mechanism, and organization of the genes. *Bioscience, Biotechnology, and Biochemistry*, 68(5):981–1003, 2004.
  - [65] J. Kaneko, T. Kimura, S. Narita, T. Tomita, and Y. Kamio. Complete nucleotide sequence and molecular characterization of the temperate staphylococcal bacteriophage  $\varphi$ pvl carrying panton–valentine leukocidin genes. *Gene*, 215(1):57–67, 1998.
  - [66] H. Karauzum, G. Chen, L. Abaandou, M. Mahmoudieh, A.R. Boroun, S. Shulenin, V.S. Devi, E. Stavale, K.L. Warfield, L. Zeitlin, et al. Synthetic human monoclonal antibodies toward staphylococcal enterotoxin b (seb) protective against toxic shock syndrome. *Journal of Biological Chemistry*, 287(30):25203–25215, 2012.
  - [67] Y. Katayama, T. Ito, and K. Hiramatsu. A new class of genetic element, staphylococcus cassette chromosome *mec*, encodes methicillin resistance in



- Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 44(6):1549–1555, 2000.
- [68] Y. Katayama, T. Ito, and K. Hiramatsu. Genetic organization of the chromosome region surrounding *mecA* in clinical staphylococcal strains: role of *IS431*-mediated *mecI* deletion in expression of resistance in textit*mecA*-carrying, low-level methicillin-resistant *Staphylococcus haemolyticus*. *Antimicrobial agents and chemotherapy*, 45(7):1955–1963, 2001.
  - [69] Y. Katayama, H.Z. Zhang, D. Hong, and H.F. Chambers. Jumping the barrier to  $\beta$ -lactam resistance in *Staphylococcus aureus*. *Journal of Bacteriology*, 185(18):5465–5472, 2003.
  - [70] H. Kikuta, M. Shibata, S. Nakata, T. Yamanaka, H. Sakata, K. Akizawa, and K. Kobayashi. Predominant dissemination of pvl-negative cc89 mrsa with *sccmec* type ii in children with impetigo in japan. *International Journal of Pediatrics*, 2011, 2011.
  - [71] E. Klein, D.L. Smith, and R. Laxminarayan. Hospitalizations and deaths caused by methicillin-resistant *Staphylococcus aureus*, united states, 1999–2005. *Emerging Infectious Diseases*, 13(12):1840, 2007.
  - [72] WE Kloos and DW Lambe. *Staphylococcus-manual of clinical microbiology*, 1991.
  - [73] J. Kluytmans and HFL Wertheim. Nasal carriage of *Staphylococcus aureus* and prevention of nosocomial infections. *Infection*, 33(1):3–8, 2005.
  - [74] V. Knoop and K. Müller. *Gene und Stammbäume*. Elsevier, Spektrum Akad. Verl., 2006.
  - [75] R. Kroker, R. Scherkl, and FR Ungemach. Chemotherapie bakterieller infektionen. *Lehrbuch der Pharmakologie und Toxikologie*. Enke Verlag, Stuttgart,, 2002.
  - [76] K. Kurt, A. Alderborn, M. Nilsson, B. Strommenger, W. Witte, and U. Nübel. Multiplexed genotyping of methicillin-resistant *Staphylococcus aureus* isolates by use of padlock probes and tag microarrays. *Journal of Clinical Microbiology*, 47(3):577, 2009.
  - [77] AR Larsen, RL Skov, V. Jarlier, and AS Henriksen. Epidemiological differences between the uk and ireland versus france in *Staphylococcus aureus* isolates resistant to fusidic acid from community-acquired skin and soft tissue infections. *Journal of Antimicrobial Chemotherapy*, 61(3):589–594, 2008.

- [78] F. Laurent, A. Tristan, M. Croze, M. Bes, H. Meugnier, G. Lina, F. Vandenesch, and J. Etienne. Presence of the epidemic european fusidic acid-resistant impetigo clone (eefic) of staphylococcus aureus in france. *Journal of Antimicrobial Chemotherapy*, 63(2):420–421, 2009.
- [79] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, et al. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.
- [80] P. Librado and J. Rozas. Dnasp v5: a software for comprehensive analysis of dna polymorphism data. *Bioinformatics*, 25(11):1451–1452, 2009.
- [81] G. Lina, G. Durand, C. Berchich, B. Short, H. Meugnier, F. Vandenesch, J. Etienne, and MC Enright. Staphylococcal chromosome cassette evolution in *Staphylococcus aureus* inferred from ccr gene complex sequence typing analysis. *Clinical microbiology and infection*, 12(12):1175–1184, 2006.
- [82] G. Lina, Y. Piémont, F. Godail-Gamot, M. Bes, M.O. Peter, V. Gauduchon, F. Vandenesch, and J. Etienne. Involvement of panton-valentine leukocidin—producing *Staphylococcus aureus* in primary skin infections and pneumonia. *Clinical Infectious Diseases*, 29(5):1128–1132, 1999.
- [83] J.A. Lindsay. Genomic variation and evolution of *Staphylococcus aureus*. *International Journal of Medical Microbiology*, 300(2-3):98–103, 2010.
- [84] J.A. Lindsay and M.T.G. Holden. *Staphylococcus aureus*: superbug, super genome? *Trends in Microbiology*, 12(8):378–385, 2004.
- [85] J.A. Lindsay and M.T.G. Holden. Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Functional & Integrative Genomics*, 6(3):186–201, 2006.
- [86] J.A. Lindsay, A. Ruzin, H.F. Ross, N. Kurepina, and R.P. Novick. The gene for toxic shock toxin is carried by a family of mobile pathogenicity islands in *Staphylococcus aureus*. *Molecular Microbiology*, 29(2):527–543, 2002.
- [87] D.M. Livermore. Antibiotic resistance in staphylococci. *International Journal of Antimicrobial Agents*, 16:3–10, 2000.
- [88] F.D. Lowy. *Staphylococcus aureus* infections. *New England Journal of Medicine*, 339(8):520–532, 1998.
- [89] X. Ma, W. Chang, C. Zhang, X. Zhou, and F. Yu. Staphylococcal panton-valentine leukocidin induces pro-inflammatory cytokine production and nuclear factor-kappa b activation in neutrophils. *PLoS One*, 7(4):e34970, 2012.

- [90] X.X. Ma, T. Ito, C. Tiensasitorn, M. Jamklang, P. Chongtrakool, S. Boyle-Vavra, R.S. Daum, and K. Hiramatsu. Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. *Antimicrobial Agents and Chemotherapy*, 46(4):1147, 2002.
- [91] M.C.J. Maiden, J.A. Bygraves, E. Feil, G. Morelli, J.E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D.A. Caugant, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, 95(6):3140–3145, 1998.
- [92] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057):376–380, 2005.
- [93] H. Masiuk, K. Kopron, D. Grumann, C. Goerke, J. Kolata, J. Jursa-Kulesza, S. Giedrys-Kalemba, B.M. Bröker, and S. Holtfreter. Association of recurrent furunculosis with panton-valentine leukocidin and the genetic background of *Staphylococcus aureus*. *Journal of Clinical Microbiology*, 48(5):1527–1535, 2010.
- [94] M.J. McGavin. Genome comparisons of diverse *Staphylococcus aureus* strains. *Bacterial Genomes and Infectious Diseases*, pages 191–212, 2006.
- [95] A. Mellmann, T. Weniger, C. Berssenbrügge, J. Rothgänger, M. Sammeth, J. Stoye, and D. Harmsen. Based upon repeat pattern (burp): an algorithm to characterize the long-term evolution of *Staphylococcus aureus* populations based on *spa* polymorphisms. *BMC Microbiology*, 7(1):98, 2007.
- [96] S. Monecke, P. Slickers, MJ Ellington, AM Kearns, and R. Ehricht. High diversity of panton–valentine leukocidin-positive, methicillin-susceptible isolates of *Staphylococcus aureus* and implications for the evolution of community-associated methicillin-resistant *S. aureus*. *Clinical Microbiology and Infection*, 13(12):1157–1164, 2007.
- [97] PCL Moore and JA Lindsay. Genetic variation among hospital isolates of methicillin-sensitive *Staphylococcus aureus*: evidence for horizontal transfer of virulence genes. *Journal of Clinical Microbiology*, 39(8):2760–2767, 2001.
- [98] KB Mullis, FA Faloona, SJ Scharf, RK Saiki, GT Horn, H. Erlich, et al. Specific enzymatic amplification of dna in vitro: the polymerase chain reaction. In *Cold Spring Harb Symp Quant Biol*, volume 51, pages 263–273, 1986.

- [99] N.E. Murray. Type i restriction systems: sophisticated molecular machines (a legacy of bertani and weigle). *Microbiology and Molecular Biology Reviews*, 64(2):412–434, 2000.
- [100] S. Muttaiyah, G. Coombs, S. Pandey, P. Reed, S. Ritchie, D. Lennon, and S. Roberts. Incidence, risk factors, and outcomes of panton-valentine leukocidin-positive methicillin-susceptible *Staphylococcus aureus* infections in auckland, new zealand. *Journal of Clinical Microbiology*, 48(10):3470–3474, 2010.
- [101] S. Narita, J. Kaneko, J. Chiba, Y. Piémont, S. Jarraud, J. Etienne, and Y. Kamio. Phage conversion of panton-valentine leukocidin in *Staphylococcus aureus*: molecular analysis of a pvl-converting phage,  $\varphi$ slt. *Gene*, 268(1):195–206, 2001.
- [102] M. Nei and T. Gojobori. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3(5):418–426, 1986.
- [103] E.K. Nickerson, V. Wuthiekanun, V. Kumar, P. Amornchai, N. Wongdeethai, K. Chheng, N. Chantratita, H. Puthhat, J. Thaipadungpanit, N.P. Day, et al. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* carriage in children in cambodia. *The American Journal of Tropical Medicine and Hygiene*, 84(2):313, 2011.
- [104] Z. Ning, A.J. Cox, and J.C. Mullikin. Ssaha: a fast search method for large dna databases. *Genome Research*, 11(10):1725–1729, 2001.
- [105] G.A. Noskin, R.J. Rubin, J.J. Schentag, J. Kluytmans, E.C. Hedblom, C. Jacobson, M. Smulders, E. Gemmen, and M. Bharmal. National trends in *Staphylococcus aureus* infection rates: impact on economic burden and mortality over a 6-year period (1998–2003). *Clinical Infectious Diseases*, 45(9):1132–1140, 2007.
- [106] M.J. Noto and G.L. Archer. A subset of *Staphylococcus aureus* strains harboring staphylococcal cassette chromosome *mec* (*sccmec*) type iv is deficient in ccrab-mediated *sccmec* excision. *Antimicrobial Agents and Chemotherapy*, 50(8):2782, 2006.
- [107] M.J. Noto, B.N. Kreiswirth, A.B. Monk, and G.L. Archer. Gene acquisition at the insertion site for *sccmec*, the genomic island conferring methicillin resistance in *Staphylococcus aureus*. *Journal of Bacteriology*, pages JB–01128, 2007.

- [108] R. Novick, A. Subedi, et al. The sapis: mobile pathogenicity islands of *Staphylococcus*. *Chemical Immunology and Allergy*, 93:42–57, 2007.
- [109] R.P. Novick. Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Molecular Microbiology*, 48(6):1429–1449, 2003.
- [110] U. Nübel, J. Dordel, K. Kurt, B. Strommenger, H. Westh, S.K. Shukla, H. Žemličková, R. Leblois, T. Wirth, T. Jombart, et al. A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant *Staphylococcus aureus*. *PLoS Pathog*, 6(4):e1000855, 2010.
- [111] U. Nübel, P. Roumagnac, M. Feldkamp, J.H. Song, K.S. Ko, Y.C. Huang, G. Coombs, M. Ip, H. Westh, R. Skov, et al. Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. *Proceedings of the National Academy of Sciences*, 105(37):14130, 2008.
- [112] A. Ogston. Report upon micro-organisms in surgical diseases. *British Medical Journal*, 1(1054):369–375, 1881.
- [113] F.P. O’Hara, N. Guex, J.M. Word, L.A. Miller, J.A. Becker, S.L. Walsh, N.E. Scangarella, J.M. West, R.M. Shawar, and H. Amrine-Madsen. A geographic variant of the *Staphylococcus aureus* panton-valentine leukocidin toxin and the origin of community-associated methicillin-resistant *S. aureus* usa300. *Journal of Infectious Diseases*, 197(2):187, 2008.
- [114] D.C. Oliveira, A. Tomasz, and H. de Lencastre. Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant *Staphylococcus aureus*. *The Lancet Infectious Diseases*, 2(3):180–189, 2002.
- [115] A. Onanuga and TC Temedie. Nasal carriage of multi-drug resistant *Staphylococcus aureus* in healthy inhabitants of amassoma in niger delta region of nigeria. *African health sciences*, 11(2), 2012.
- [116] AJ O’Neill, AR Larsen, R. Skov, AS Henriksen, and I. Chopra. Characterization of the epidemic european fusidic acid-resistant impetigo clone of *Staphylococcus aureus*. *Journal of Clinical Microbiology*, 45(5):1505, 2007.
- [117] M. Otto. Basis of virulence in community-associated methicillin-resistant staphylococcus aureus\*. *Annual review of microbiology*, 64:143–162, 2010.
- [118] M. Otto, R. Süßmuth, C. Vuong, G. Jung, and F. Götz. Inhibition of virulence factor expression in *Staphylococcus aureus* by the *Staphylococcus epidermidis* agr pheromone and derivatives. *FEBS letters*, 450(3):257–262, 1999.
- [119] PN Panton, MB Came, and FCO Valentine. Staphylococcal toxin. *The Lancet*, 1932.

- [120] R. Pantucek, J. Doskar, V. Ruzickova, P. Kasperek, E. Oracova, V. Kvardova, and S. Rosypal. Identification of bacteriophage types and their carriage in *Staphylococcus aureus*. *Archives of virology*, 149(9):1689–1703, 2004.
- [121] J. Parker, A. Rambaut, and O.G. Pybus. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infection, Genetics and Evolution*, 8(3):239–246, 2008.
- [122] A.S. Payne, Y. Hanakawa, M. Amagai, and J.R. Stanley. Desmosomes and disease: pemphigus and bullous impetigo. *Current Opinion in Cell Biology*, 16(5):536–543, 2004.
- [123] S.J. Peacock, I. de Silva, and F.D. Lowy. What determines nasal carriage of *Staphylococcus aureus*? *Trends in Microbiology*, 9(12):605–610, 2001.
- [124] T. Pearson, J.D. Busch, J. Ravel, T.D. Read, S.D. Rhoton, J.M. U'ren, T.S. Simonson, S.M. Kachur, R.R. Leadem, M.L. Cardon, et al. Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(37):13536–13541, 2004.
- [125] M.G. Pinho, H. De Lencastre, and A. Tomasz. An acquired and a native penicillin-binding protein cooperate in building the cell wall of drug-resistant staphylococci. *Proceedings of the National Academy of Sciences*, 98(19):10886–10891, 2001.
- [126] J.P. Rasigade, F. Laurent, G. Lina, H. Meugnier, M. Bes, F. Vandenesch, J. Etienne, and A. Tristan. Global distribution and evolution of panton-valentine leukocidin-positive methicillin-susceptible *Staphylococcus aureus*, 1981–2007. *Journal of Infectious Diseases*, 201(10):1589, 2010.
- [127] MIA Rijnders, PFG Wolffs, RM Hopstaken, M. den Heyer, CA Bruggeman, and EE Stobberingh. Spread of the epidemic european fusidic acid-resistant impetigo clone (eefic) in general practice patients in the south of the netherlands. *Journal of Antimicrobial Chemotherapy*, 2012.
- [128] G. Ritter von Rittershain. 1 878. die ex foliative dermatitis jiingeren siiuglinge. *Z. Kinderheilkd*, 2:3–23.
- [129] D.A. Robinson and M.C. Enright. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy*, 47(12):3926–3934, 2003.
- [130] M. Ronaghi, M. Uhlén, and P. Nyrén. A sequencing method based on real-time pyrophosphate. *Science*, 281(5375):363–365, 1998.

- [131] P. Roumagnac, F.X. Weill, C. Dolecek, S. Baker, S. Brisse, N.T. Chinh, T.A.H. Le, C.J. Acosta, J. Farrar, G. Dougan, et al. Evolutionary history of *Salmonella typhi*. *Science*, 314(5803):1301–1304, 2006.
- [132] R. Ruimy, A. Maiga, L. Armand-Lefevre, I. Maiga, A. Diallo, A.K. Koumaré, K. Ouattara, S. Soumaré, K. Gaillard, J.C. Lucet, et al. The carriage population of *Staphylococcus aureus* from mali is composed of a combination of pandemic clones and the divergent panton-valentine leukocidin-positive genotype st152. *Journal of Bacteriology*, 190(11):3962–3968, 2008.
- [133] V. Růžicková, R. Pantůček, P. Petráš, I. Machová, K. Kostýlková, and J. Doškař. Major clonal lineages in impetigo *Staphylococcus aureus* strains isolated in czech and slovak maternity hospitals. *International Journal of Medical Microbiology*, 2012.
- [134] A. Ruzin, J. Lindsay, and R.P. Novick. Molecular genetics of sapi1—a mobile pathogenicity island in *Staphylococcus aureus*. *Molecular Microbiology*, 41(2):365–377, 2001.
- [135] M. Salemi, S.L. Lamers, S. Yu, T. De Oliveira, W.M. Fitch, and M.S. McGrath. Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of aids. *Journal of Virology*, 79(17):11343–11352, 2005.
- [136] F. Sanger, S. Nicklen, and A.R. Coulson. Dna sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463, 1977.
- [137] D.I. Scaduto, J.M. Brown, W.C. Haaland, D.J. Zwickl, D.M. Hillis, and M.L. Metzker. Source identification in two criminal cases using phylogenetic analysis of hiv-1 dna sequences. *Proceedings of the National Academy of Sciences*, 107(50):21242–21247, 2010.
- [138] D.C. Schwartz, C.R. Cantor, et al. Separation of yeast chromosome-sized dnas by pulsed field gradient gel electrophoresis. *Cell*, 37(1):67, 1984.
- [139] C.B. Smith, V. Noble, R. Bensch, P.A. Ahlin, JA Jacobson, R.H. Latham, et al. Bacterial flora of the vagina during the menstrual cycle: findings in users of tampons, napkins, and sea sponges. *Annals of internal medicine*, 96(6 Pt 2):948, 1982.
- [140] M.E. Stemper, S.K. Shukla, and K.D. Reed. Emergence and spread of community-associated methicillin-resistant *Staphylococcus aureus* in rural wisconsin, 1989 to 1999. *Journal of Clinical Microbiology*, 42(12):5673–5680, 2004.

- [141] B. Strommenger, C. Bräulke, D. Heuck, C. Schmidt, B. Pasemann, U. Nübel, and W. Witte. *spa* typing of *Staphylococcus aureus* as a frontline tool in epidemiological typing. *Journal of Clinical Microbiology*, 46(2):574–581, 2008.
- [142] S.Y.C. Tong, R.A. Lilliebridge, E.J. Bishop, A.C. Cheng, D.C. Holt, M.I. McDonald, P.M. Giffard, B.J. Currie, and C.S. Boutlis. Clinical correlates of panton-valentine leukocidin (pvl), pvl isoforms, and clonal complex in the *Staphylococcus aureus* population of northern australia. *Journal of Infectious Diseases*, 202(5):760–769, 2010.
- [143] T. Tsuru and I. Kobayashi. Multiple genome comparison within a bacterial species reveals a unit of evolution spanning two adjacent genes in a tandem paralog cluster. *Molecular Biology and Evolution*, 25(11):2457–2473, 2008.
- [144] K.M.E. Turner and E.J. Feil. The secret life of the multilocus sequence type. *International Journal of Antimicrobial Agents*, 29(2):129–135, 2007.
- [145] J.D. Turnidge, JM Bell, et al. Methicillin-resistant *Staphylococcal aureus* evolution in australia over 35 years. *Microbial Drug Resistance*, 6(3):223–229, 2000.
- [146] J. Valle, S. Piriz, R. Fuente, and S. Vadillo. Staphylococci isolated from healthy goats. *Journal of Veterinary Medicine, Series B*, 38(1-10):81–89, 1991.
- [147] H. Van de Velde. Etude sur le mécanisme de la virulence du staphylocoque pyogène. *La Cellule*, 10:401–410, 1894.
- [148] G.M. Vath, C.A. Earhart, D.D. Monie, J.J. Iandolo, P.M. Schlievert, and D.H. Ohlendorf. The crystal structure of exfoliative toxin b: a superantigen with enzymatic activity. *Biochemistry*, 38(32):10239–10246, 1999.
- [149] G.M. Vath, C.A. Earhart, J.V. Rago, M.H. Kim, G.A. Bohach, P.M. Schlievert, and D.H. Ohlendorf. The structure of the superantigen exfoliative toxin a suggests a novel regulation as a serine protease. *Biochemistry*, 36(7):1559–1566, 1997.
- [150] H. Veiga and M.G. Pinho. Inactivation of the sauI type I restriction-modification system is not sufficient to generate *Staphylococcus aureus* strains capable of efficiently accepting foreign dna. *Applied and environmental microbiology*, 75(10):3034–3038, 2009.
- [151] V. Vorobieva, T. Bazhukova, A.M. Hanssen, DA Caugant, N. Semenova, B.C. Haldorsen, G.S. Simonsen, and A. Sundsfjord. Clinical isolates of *Staphylococcus aureus* from the arkhangelsk region, russia: antimicrobial susceptibility, molecular epidemiology, and distribution of panton-valentine leukocidin genes. *Apmis*, 116(10):877–887, 2008.



- [152] D.E. Waldron and J.A. Lindsay. Saul: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. *Journal of Bacteriology*, 188(15):5578–5585, 2006.
- [153] TH Wang, YK Donaldson, RP Brette, JE Bell, and P. Simmonds. Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *Journal of Virology*, 75(23):11686–11699, 2001.
- [154] R. Warren, M. Rogolsky, B.B. Wiley, and L.A. Glasgow. Effect of ethidium bromide on elimination of exfoliative toxin and bacteriocin production in *Staphylococcus aureus*. *Journal of Bacteriology*, 118(3):980–985, 1974.
- [155] M. Whitby, M.L. McLaws, G. Berry, et al. Risk of death from methicillin-resistant *Staphylococcus aureus* bacteraemia: a meta-analysis. *Medical Journal of Australia*, 175(5):264–267, 2001.
- [156] J.C. Wilgenbusch and D. Swofford. Inferring evolutionary trees with paup\*. *Current Protocols in Bioinformatics*, 2003.
- [157] REO Williams. Healthy carriage of *Staphylococcus aureus*: its prevalence and importance. *Bacteriological Reviews*, 27(1):56, 1963.
- [158] C. Wirtz, W. Witte, C. Wolz, and C. Goerke. Insertion of host dna into pvl-encoding phages of the *Staphylococcus aureus* lineage st80 by intra-chromosomal recombination. *Virology*, 406(2):322–327, 2010.
- [159] W. Witte, C. Cuny, I. Klare, U. Nübel, B. Strommenger, and G. Werner. Emergence and spread of antibiotic-resistant gram-positive bacterial pathogens. *International Journal of Medical Microbiology*, 298(5-6):365–377, 2008.
- [160] S. Wu, C. Piscitelli, H. de Lencastre, and A. Tomasz. Tracking the evolutionary origin of the methicillin resistance gene: cloning and sequencing of a homologue of *mecA* from a methicillin susceptible strain of *Staphylococcus sciuri*. *Microbial Drug Resistance*, 2(4):435–441, 1996.
- [161] T. Yamaguchi, T. Hayashi, H. Takami, K. Nakasone, M. Ohnishi, K. Nakayama, S. Yamada, H. Komatsuzawa, and M. Sugai. Phage conversion of exfoliative toxin a production in *Staphylococcus aureus*. *Molecular Microbiology*, 38(4):694–705, 2000.
- [162] T. Yamaguchi, T. Hayashi, H. Takami, M. Ohnishi, T. Murata, K. Nakayama, K. Asakawa, M. Ohara, H. Komatsuzawa, and M. Sugai. Complete nucleotide

- sequence of a *Staphylococcus aureus* exfoliative toxin b plasmid and identification of a novel adp-ribosyltransferase, edin-c. *Infection and Immunity*, 69(12):7760–7771, 2001.
- [163] J. Zhang, R. Chiodini, A. Badr, and G. Zhang. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 2011.

## Danksagung

Zu Beginn möchte ich mich bei Herrn Dr. Ulrich Nübel und Prof. Dr. Wolfgang Witte für die Möglichkeit und für das entgegengebrachte Vertrauen bedanken. Ohne die die Bearbeitung des Themas am Robert Koch-Institut nicht möglich gewesen wäre. Besonderer Dank gilt dabei Dr. Ulrich Nübel der mir in den vielen Jahren der Zusammenarbeit, die sowohl meine Studienarbeit als auch meine Diplomarbeit umfassen, eine hervorragende Begleitung und wissenschaftliche Ausbildung ermöglichte und schließlich auch mein Mentor wurde. Durch die Chance mich auf nationalen und internationalen Tagungen beweisen zu können und im Rahmen einer Kooperation auch zeitweise in Irland zu arbeiten, möchte ich dir ebenfalls noch sehr danken.

Bei Prof. Dr. Michael Steinert bedanke ich mich für die unkomplizierte Übernahme meiner Dissertation als Zweitgutachter.

Bei Prof. Dr. Andre Fleißner für die Bereitschaft als Fachprüfer der Promotionskommission beizusitzen und den Vorsitz zu übernehmen.

Bei Annette Weller, Mike Henkel, Franziska Erdmann, Heike Illiger, Wiebke Streckel, Christa Cuny möchte ich mich für die stetige Hilfe und Unterstützung im Labor bedanken, ohne die diese Arbeit nicht möglich gewesen wäre. Des Weiteren möchte ich mich auch bei allen Kollegen und Kolleginnen des Robert Koch-Institutes sowohl in Berlin, als auch in Wernigerode, insbesondere FG13, für die freundliche Aufnahme und die entgegengebrachte Hilfe.

Für die außerordentlich gute, kreative und alltagsvergessende Stimmung möchte ich mich ganz speziell bei meinen sowohl ehemaligen als auch noch derzeitigen Kollegen in Wernigerode bedanken. Dies gilt insbesondere für Jennycita, Jenny „Äffchen“, Ulli, Anne Wittenberg, Christoph „Elli“, Mohamed „Mojo,Mo“, Katja „Coolchen“, Thomas, Maren, Christina, Franzi, Birgit, Guido, Axenia „Axi“, Tine „rotes Eichhörnchen“, Magda „schwarzes Eichhörnchen“, Janina, Matthias „Mathi“, Thorsten, Kathrin, Tobi K. und Franzi Faber. Ich danke euch für die vielen Gespräche, Diskussionen, Flusen, Spiele und abwechslungsreichen Nachmittage und bin froh euch kennengelernt zu haben.

Meiner Familie, besonders meinen Eltern Gabriele Spamer und Hartmut Kurt sowie meinem Stiefvater Christoph Spamer, sei an dieser Stelle ebenfalls sehr gedankt. Sie bildeten ein festes Fundament für meine persönliche Entwicklung und haben mir immer viel Liebe, Vertrauen und Unterstützung entgegenbracht.